SWITCHED LINEAR SYSTEMS: OBSERVABILITY AND OBSERVERS

A Thesis
Presented to
The Academic Faculty

by

Mohamed Babaali

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
March, 2004

SWITCHED LINEAR SYSTEMS: OBSERVABILITY AND OBSERVERS

Approved by:

Professor Magnus Egerstedt, Advisor

Professor Erik Verriest

Professor Yorai Wardi

Date Approved: March 22, 2004

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my thesis advisor, Professor Magnus Egerstedt, for his guidance, patience, and support. I am also specially grateful to Professor Edward W. Kamen for his support and encouragement early on in my PhD studies, as my first advisor. I also wish to thank Professors Erik I. Verriest and Yorai Wardi for serving on my reading committee, and Professors Anthony Yezzi and Yang Wang for being on my defense committee.

It is a pleasure to thank my fellow graduate students for their friendship, their advices, and many stimulating conversations: Mohamad Abou El-Nasr, Adam Austin, Henrik Axelsson, Mauro Boccadoro, Florent Delmotte, Meng Ji, Tejas Mehta, Jeongseung Moon, and Abubakr Muhammad.

Finally, I cannot overstate my gratitude to my family for their endless support in all of my endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Switched linear systems have long been subject to high interest and intense research efforts, not only because many real world systems happen to exhibit switching behaviors, but also because the control of many complex systems is only possible via the combination of classical continuous control laws with supervisory switching logic.

A particularly important problem is that of estimator and observer design, since the state of a system is usually only available through partial, often noise-corrupted, measurements. Even though hybrid estimation has been around for at least thirty years, a veil of mystery has surrounded the concept of "observability" in switched linear systems. It is not until recently, with the recent renewal of interest toward deterministic hybrid systems, that observer design and observability analysis have fuelled sustained research efforts. It is in this context that this work is grounded. More precisely, the objective of this research is twofold:

- To define proper concepts of observability in discrete-time switched linear systems, to characterize them, and to analyze their main properties, among which decidability is of special importance.

- To propose and analyze observers - deadbeat and asymptotic - for such systems.

The main contributions of this dissertation are as follows. It is shown that pathwise observability, i.e. state observability under arbitrary mode sequences, is decidable. Furthermore, the Kalman-Bertram sampling criterion is carried over to switched linear systems. Under unknown modes, mode and state observability are both characterized through simple linear algebraic tests, and are shown to be decidable in the autonomous case. As for asymptotic observers, a direct algebraic approach is analyzed for the class of linear systems subjected to switching in the measurement equation.

CHAPTER 1

INTRODUCTION

1.1 Motivation

First, let us motivate both the model and the problems under consideration.

1.1.1 Model

The general model under consideration here is:

$$\begin{aligned} x_{k+1} &= A(\theta_k)x_k + B(\theta_k)u_k \\ y_k &= C(\theta_k)x_k, \end{aligned} \tag{1}$$

where x_k (the state), u_k (the control), and y_k (the observation, or measurement) are in $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^p$ respectively, and where $A(\cdot)$, $B(\cdot)$, and $C(\cdot)$, are matrices of compatible dimensions. θ_k, which will be referred to as the mode in effect at time k, takes values in some finite set $\{1, \ldots, s\}$, indexing the parameter matrices in such a way that they switch among s different values. Hence the denomination *switched linear system*, which will be abbreviated as SLS throughout this dissertation.

Moreover, as opposed to hybrid systems (e.g., piecewise linear systems), we characterize switched linear systems by the fact the modes are independent of all other variables. In other words, we assume θ_k to be an exogenous variable that is governed by some external process, which could be the controller itself. Finally, we will assume throughout that the possible mode sequences $\{\theta_k\}_{k=1}^{\infty}$ are arbitrary. One important consequence of such an assumption is the absence of a minimum time separation between consecutive switches.

Switched systems arise in many contexts. In fact, the SLS model (1) provides such a general framework that one simply cannot enumerate an exhaustive list of situations falling under its umbrella. Instead, we mention the main problems that have triggered and fuelled research on SLS's:

- Plant failures: When a plant undergoes random failures, it is possible to model a finite number of failure modes with the s different modes in (1). For example, sensor or actuator failures, when occurring in a finite number of known failure modes, call for different $C(\cdot)$ or $B(\cdot)$ matrices, respectively.

- Target tracking: the hybrid model (1) can be used to capture the different maneuvers of maneuvering targets, as well as the data association problem.

1.1.2 Problem Formulation

In order to control the SLS (1), the state x_k is usually required for feedback, while only the observations y_k are available. This calls for the determination of the states from the observations. An algorithm used to compute the state from the measurements is termed an *observer* [53]. This is the problem we are concerned with in this dissertation. We will distinguish between two kinds of observers:

- Deadbeat observers, which determine x_1 from a finite number of measurements $y_1, \ldots, y_N$.

- Asymptotic observers, which compute an estimate $\hat{x}_k$ of the state x_k from $y_1, \ldots, y_k$, such that

$$\lim_{k\to\infty} \|x_k - \hat{x}_k\| = 0, \tag{2}$$

  where $\|\cdot\|$ is some norm in $\mathbb{R}^n$.

The conditions under which there exists a deadbeat observer will in general be referred to as *observability*, following Kalman's terminology [42]. Let us recall the main results on observability for standard linear systems:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k \end{aligned} \tag{3}$$

Definition 1 (Observability [42]) *The linear system (3) is observable if there exists a time horizon N such that the first N observations $y_1, \ldots, y_N$ determine the initial state x_1 uniquely.* $\diamondsuit$

Noting that we can restrict our attention to the autonomous case without loss of generality, and that the observations can then be written as follows:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \underbrace{\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{pmatrix}}_{\mathcal{O}_N} x_1, \tag{4}$$

where $\mathcal{O}_N$ is the Kalman *observability matrix*, we realize that the observations are nothing but a linear function of the initial state. Therefore, the matrix of that transformation, which is $\mathcal{O}_N$, must be of full column rank n. In other words:

Theorem 1 *The linear system (3) is observable if and only if there exists an integer N such that $\rho(\mathcal{O}_N) = n$, or equivalently if and only if $\rho(\mathcal{O}_n) = n$.* $\diamondsuit$

Note that, while the first condition in Theorem 1, i.e. the existence of N such that $\rho(\mathcal{O}_N) = n$, does not call for an algorithm, the second condition reduces it to a single matrix rank test, which can easily be computed, making the observability of a linear system *decidable*. One of the objectives of this thesis is to carry this analysis over to SLS's. More precisely, whenever possible, attempts will be made to:

- Translate observability problems into linear algebraic tests.
- To assess their decidability.

Finally, note that the addition of switching to the linear system (3) will not only increase the complexity of the observability analysis and observer design problems, but will also multiply the number of problems for consideration. Indeed, we will need to distinguish between:

- The known and the unknown modes cases: Depending on the application, the modes θ_k are either known, i.e. available to the observer, or unknown, making the state observability problem more complex.

- State and mode observability: In the unknown modes case, one may wish to recover either the modes or the states.

- The autonomous and the non-autonomous cases: While the separation principle for linear time-varying systems allows one to decouple observation from control in the known modes case, it turns out that the controls have an effect on the various observability concepts in the unknown modes case. This will call for the distinction between the autonomous and the non-autonomous cases, and in the latter case, for the consideration of *single experiment* problems and *generic experiment* problems.

1.2 Background

Discrete-time SLS's lie among numerous other classes of multi-modal linear systems, a rough classification of which can be given in terms of the following attributes:

- Deterministic vs. stochastic systems.

- Discrete-time vs. continuous-time systems.

- Switched vs. hybrid systems.

Virtually every combination of these attributes has been considered in the literature,[1] and requires an analysis of its own. For instance, while observability analysis in

[1] Historically, it appears that the the stochastic class of Markov jump linear systems in continuous-time was the first one to be studied, from an optimal control point of view [52, 67].

discrete-time and continuous-time linear systems is virtually the same, significant differences arise in SLS, as we will later point out. Although an exhaustive survey of observability analysis and observer design for all of these models is beyond our scope, we will describe the main results concerning close models in the next subsection. Then, in the following subsection, we will describe the background of our specific problem, i.e. observability analysis and observer design for discrete-time switched linear systems.

1.2.1 Related Problems

In this subsection, we will briefly discuss the main results from hybrid estimation in stochastic switched linear systems, from observability analysis in hybrid systems, and finally from observability analysis in continuous-time switched linear systems.

1.2.1.1 Hybrid Estimation

The most important class of stochastic switched linear systems can be described by the following model:

$$\begin{aligned} x_{k+1} &= A(\theta_k)x_k + B(\theta_k)u_k + w_k \\ y_k &= C(\theta_k)x_k + v_k, \end{aligned} \tag{5}$$

where w_k and v_k are white Gaussian noises, while all other variables and parameters are as in (1). θ_k, however, is assumed unknown, and is endowed with a probabilistic model. In most cases, the modes θ_k have been assumed to be governed by a homogeneous Markov chain. Such systems are usually referred to as *jump Markov linear systems* (JMLS).

The field of hybrid estimation, i.e. optimal estimator design for (5), has been very active since its inception, which, under a general consensus, occurred with the 1970 paper by Ackerson and Fu [1], who provided the first derivation of the minimum mean-squared error (MMSE) estimator for (5). The MMSE estimator being impossible to implement in practice because of exponentially growing computational requirements with time, research was quickly directed toward the design of computationally efficient sub-optimal filters. Furthermore, the desperate need of the target tracking community for efficient estimators provided extra motivation for such lines of research. While the first sub-optimal practical estimator was also proposed in [1], intense activity ensued in hybrid estimation, as testified to by the numerous surveys [49, 59, 68] and books [13] published. The main breakthroughs have been Blom's celebrated interacting multiple model (IMM) filter [16], which, until very recently, was considered to be the most efficient hybrid estimator, and particle filtering [28]. Note that much less effort was devoted to the continuous-time case, since the results were mainly motivated by practical digital implementations (see [56] for an exhaustive survey up to 1990).

It is noteworthy that convergence analysis was virtually absent from all of this work, which is due to the complexity of hybrid estimators introduced by the explicit computation of gaussian probabilities. In particular, no convergence theorem exists today for any estimator for (5). In fact, even the Kalman filter had to wait for 8 years, i.e. for the results in [27], for its first stability (actually, asymptotic unbiasedness) result. Finally, even though observability is actually an inherently deterministic concept, it is noteworthy that one can find numerous papers on observability analysis for JMLS [24, 37, 38, 55, 73]. It turns out that those papers actually study convergence problems, rather than observability *per se*.

1.2.1.2 Hybrid Systems

We now return to the deterministic case (1), except that we assume that the modes θ_k are no longer independent of the other variables. Such systems, where continuous-valued variables interact with discrete-valued ones, are termed *hybrid* in the literature.

The simplest class of discrete-time linear hybrid systems is that of *piecewise affine systems* (PWAS), where θ_k is is a piecewise constant function of (x_k, u_k). This class was first studied by Sontag in [64], where it was proposed as a way of approximating nonlinear systems. Unfortunately, observability, under its natural definition, was shown there to be undecidable, and checking whether the index of observability is smaller than a certain integer was then shown to be $\mathcal{NP}$-complete in [65]. In the nineties, later results on neural nets by the same author resulted in proofs that observability in even simpler classes of piecewise affine systems, such as systems with saturated outputs, was undecidable [61, 66].

The curious fact to note here is that Sontag's early negative results did not permanently kill all observability analysis efforts for PWAS's. In fact, along with the recognition of hybrid systems as an important area of future research in the nineties [32, 4], the recent years have witnessed renewed effort in observability analysis for hybrid systems, especially on the computational front. The main line of work to mention here is that based on the *mixed logical dynamical* (MLD) formulation of PWAS's [15, 14, 31], which has allowed for efficient observability tests to be proposed, based on mixed-integer linear programming (MILP) and on multi-parametric MILP programming. The same frenzy has lately also reached continuous-time hybrid systems [12, 23, 26, 72].

1.2.1.3 Continuous-Time Switched Linear systems

The only analysis of observability for continuous-time SLS's can be found in the early work of Ezzine and Haddad in [30], where they considered the special class of systems with periodic switching.

Note that, in continuous-time, *instantaneous observability* of a SLS, which is defined as the ability to recover the initial state uniquely from the derivatives of the output under known modes, reduces to observability of each mode. Therefore, decidability is not an issue, as opposed to the discrete-time case.

1.2.2 Switched Linear Systems

While stability and controllability analysis for the SLS (1) have mobilized considerable effort since the 1980's, as summarized in [17, 51],[2] it was not until the turn of the century that the first serious research on observability analysis [71] and on asymptotic observer design [3, 11] was published.

In [71], Vidal et al. carried out the first systematic attempt at characterizing observability concepts in SLS's. However, their work not only restricts the switching by imposing an unnecessary minimum separation between consecutive switches, it also contains several flaws that are corrected in Chapter 4. The main point that was missed in [71] is that mode observability depends on the initial state, which furthermore calls for mode and state observability to be studied independently.

In [3], Alessandri and Coletta described an asymptotic observer design approach for SLS's with known modes. More precisely, they showed how to find Luenberger gains $K(\cdot)$ such that the following

$$\hat{x}_{k+1} = A(\theta_k)\hat{x}_k + B(\theta_k)u_k + K(\theta_k)\left(y_k - C(\theta_k)\hat{x}_k\right) \tag{6}$$

results in an exponential observer for (1). Their method was based on the use of linear matrix inequalities (LMI's) to compute gains such that the dynamics of the observer error

$$e_{k+1} = \left(A(\theta_k) - K(\theta_k)C(\theta_k)\right)e_k \tag{7}$$

admit a common quadratic Lyapunov function. Balluchi et al. [11] then extended this approach to the unknown modes case, by borrowing tools from the failure detection literature. However, since residual-based failure detection is usually delayed, it requires slow switching, which, in turn, imposes slow switching for the observer to

[2]The most remarkable recent advances being the proof of stability under solvability of the Lie group generated by the A matrices [50], the characterization of controllability [33], and the disproof of the Finiteness Conjecture [18], which had been previously emitted and studied in [25, 36, 48]. Note that the decidability of stability under arbitrary switching is still an open problem.

function. Moreover, because of the delayed detection, all that could be guaranteed in [11] was an upper bound on the norm of the observer error.

1.3 Organization

This dissertation is organized as follows.

In Chapter 2, it is shown that *pathwise observability*, which is observability of SLS's under known modes, is decidable. Further results on pathwise observability are provided in Chapter 3, where the Kalman Bertram criterion for the conservation of observability after sampling is carried over to the class of SLS's undergoing switching in only the measurement equation.

In Chapter 4, several observability concepts for SLS's under unknown modes, including discernibility, are defined, characterized, and shown to be decidable in the autonomous case.

In Chapter 5, the Direct Algebraic Approach (DAA), a novel asymptotic observer design approach for a class of SLS's, is described, and its stability analyzed.

Finally, in Chapter 6, the contributions are summarized, and some future work is suggested.

CHAPTER 2

PATHWISE OBSERVABILITY: DECIDABILITY

2.1 Introduction

Let us recall our model for SLS's:

$$\begin{aligned} x_{k+1} &= A(\theta_k)x_k + B(\theta_k)u_k,\ k \geq 1 \\ y_k &= C(\theta_k)x_k, \end{aligned} \tag{8}$$

where x_k, u_k and y_k are in $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^p$ respectively, and where $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ are real matrices of compatible dimensions. The mode θ_k assumes values in the set $\{1, \dots, s\}$, so that the parameter matrices switch between s different known values. Again, we furthermore assume that the mode sequence $\{\theta_k\}_{k=1}^{\infty}$ is arbitrary and independent of the initial state x_1 and input sequence $\{u_k\}_{k=1}^{\infty}$. In this chapter, we examine a property of (8) concerning both observability and controllability, that can be motivated as follows: It is clear that given y_k, u_k and θ_k, $k = 1, \dots, N$, it is possible to recover x_1 uniquely if and only if the following observability matrix has full column rank n:

$$\begin{pmatrix} C(\theta_1) \\ C(\theta_2)A(\theta_1) \\ \vdots \\ C(\theta_N)A(\theta_{N-1}) \cdots A(\theta_1) \end{pmatrix}, \tag{9}$$

in which case x_1 is uniquely given by

$$x_1 = M \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \tag{10}$$

where M is the $\{1\}$-inverse of the matrix in (9) (see Appendix A). If we further want to be able to recover x_1 for any sequence $\theta_1, \ldots, \theta_N$, then all such matrices must be of full rank. We define the property wherein such an N exists as *pathwise observability*, and we note that it has appeared in the literature for quite some time, starting with [38], where it was linked to a concept of stochastic observability, and where *pathwise controllability* was also considered and linked to the existence of steady-state solutions to the Markov jump linear quadratic problem. More recently, it was shown in [20, 44] that pathwise observability implied the existence of artificial stochastic parameters such that the corresponding Kalman filter results in an asymptotic observer for (8), and, in [9], an asymptotic observer was proposed for a special subclass of (8), whose convergence was established under similar assumptions. However, what has been missing is a way to check for pathwise observability. The direct way is to check the rank of all matrices (9) for increasing N until they all reach full rank, or until it is provably impossible for pathwise observability to hold. However, while it is well known [40], thanks to the Cayley-Hamilton Theorem, that this algorithm terminates at $N = n$ for standard (unimodal) linear systems, it has been unknown whether or not it terminates for switched linear systems. In this chapter, we affirmatively answer this question, and we provide finite upper bounds on the maximum *index* of pathwise observability. Notice that these results imply that pathwise observability is decidable, and, as it turns out, by duality, that pathwise controllability is decidable as well.

This chapter is organized as follows: We start off, in Section 2.2, by establishing some definitions and by stating the main results. We devote Section 2.3 to the proofs. Finally, we study the dual problem of pathwise controllability in Section 2.4.

2.2 Definitions and Results

Without loss of generality, we restrict our analysis to autonomous switched linear systems of the form:

$$x_{k+1} = A(\theta_k)x_k,$$
$$y_k = C(\theta_k)x_k,$$

which we characterize by the set of pairs $\{(A(1), C(1)), \ldots, (A(s), C(s))\}$. We define a path θ of length N as a string $\theta_1\theta_2\cdots\theta_N$ over the set $\{1, \ldots, s\}$. We let the length of such a string be denoted by $|\theta| = N$. We also define the observability matrix of a path θ of length N as:

$$\mathcal{O}(\theta) \triangleq \begin{pmatrix} C(\theta_1) \\ C(\theta_2)A(\theta_1) \\ \vdots \\ C(\theta_N)A(\theta_{N-1})\cdots A(\theta_1) \end{pmatrix}.$$

We furthermore say that a path θ has rank r if and only if its observability matrix $\mathcal{O}(\theta)$ has rank r. Similarly, a path is observable if and only if its observability matrix has full rank n. Finally, we have the following definition:

Definition 2 (Pathwise Observability) *The set of pairs* $\{(A(1), C(1)), \ldots, (A(s), C(s))\}$ *is pathwise observable if and only if there exists an integer N such that all paths of length N are observable. We refer to the smallest such integer as the index of pathwise observability.* $\diamondsuit$

If a set of pairs is not pathwise observable (i.e. for all N, there exists an unobservable path of length N), it is said to be pathwise unobservable.

Remarks 1 *The following trivial remarks can now be made:*

- *A set of pairs containing an unobservable pair is pathwise unobservable.*

- *A set of pairs may contain only observable pairs, and yet be path-wise unobservable. For example, consider* $\{(A(1), C(1)), (A(2), C(2))\}$*, where*

$$A(1) = A(2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \begin{aligned} C(1) &= (1 \quad 0) \\ C(2) &= (0 \quad 1) \end{aligned} \tag{11}$$

- *A set of pairs with only unobservable pairs (thus pathwise unobservable) may have observable paths. For example, consider* $\{(A(1), C(1)), (A(2), C(2))\}$*, where*

$$A(1) = A(2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{aligned} C(1) &= (1 \quad 0) \\ C(2) &= (0 \quad 1) \end{aligned} \tag{12}$$

- *Unlike with linear time invariant systems, observability of paths may be reached at arbitrary path lengths. For example, consider the previous example, with the following paths:*

$$\theta_k = 1, \qquad k = 1, \ldots, N \tag{13}$$

$$\theta_N = 2. \tag{14}$$

  Every one of these paths has no observable prefix, and therefore becomes observable at N*, which can be chosen arbitrarily large.*

We moreover need to define the pathwise r-rank property as follows:

Definition 3 (Pathwise r-rank) *The set of pairs* $\{(A(1), C(1)), \ldots, (A(s), C(s))\}$ *is pathwise* r*-ranked if and only if there exists an integer* N *such that the rank of every path of length* N *is at least* r*. We refer to the smallest such integer as the index of pathwise* r*-rank.* $\diamondsuit$

Note that the pathwise n-rank property is equivalent to pathwise observability. In this chapter, we show that the pathwise r-rank property is decidable for all $r \leq n$, which includes pathwise observability. Now, before stating the main result of this chapter, we need to define the following quantities:

$$
\begin{aligned}
N(s,n,1) &\triangleq 1 \\
N(s,n,r) &\triangleq G(r, N(s,n,r-1), s^{N(s,n,r-1)}, r), \ r \leq n,
\end{aligned}
$$

where $G(r,g,p,k)$ is computed recursively as follows, by a double induction on k and p:

$$
\begin{aligned}
G(r,g,1,r) &\triangleq l+1 \\
G(r,g,p,k+1) &\triangleq G(k+1, G(r,g,p,k), s^{G(r,g,p,k)}, k+1), \ k = r, \ldots, n-1, \\
G(r,g,p+1,r) &\triangleq 1 + \max_{k=r,r+1,\ldots,n} \{G(r,g,p,k)\}, \ p = 1, \ldots, s^g - 1.
\end{aligned}
$$

We also define:

$$
\begin{aligned}
N_c(s,n,1) &\triangleq 1, \\
N_c(s,n,r) &\triangleq N_c(s,n,r-1) + s^{N_c(s,n,r-1)}, \ 2 \leq r \leq n.
\end{aligned}
$$

We can now state the following theorem:

Theorem 2 *Assume given a set of s pairs $\{(A(1),C(1)),\ldots,(A(s),C(s))\}$, where the dimension of the $A(\cdot)$ matrices is n.*

1. *If the set is pathwise r-ranked, then its index of pathwise r-rank is smaller than or equal to $N(s,n,r)$.*

2. *If furthermore the $A(\cdot)$ matrices are pairwise commuting, then the index of pathwise r-rank is bounded by the smaller number $N_c(s,n,r)$.* $\Diamond$

The proof is given in Section 2.3, and a corollary to Theorem 2 reads as follows:

Corollary 1 *The pathwise r-rank and pathwise observability properties are decidable.* $\Diamond$

Table 1: Values of $N(s, n, n)$ for small n and s

n \ s	1	2	3
1	1	1	1
2	2	3	4
3	3	135	?

Proof: It clearly suffices to compute the rank of $\mathcal{O}(\theta)$ for every path θ of length $N(s, n, r)$ (resp. $N(s, n, n)$). A set of pairs is then pathwise r-ranked (resp. observable) if and only if the rank of $\mathcal{O}(\theta)$ for every such θ is at least r (resp. equals n). $\square$

We now define $\mathcal{N}(s, n, r)$ as the maximum index of pathwise r-rank over all pathwise r-ranked sets of pairs $\{(A(1), C(1)), \ldots, (A(s), C(s))\}$ with $n \times n$ $A(\cdot)$ matrices. Similarly, let $\mathcal{N}_c(s, n, r)$ be the maximum index of pathwise r-rank over all pathwise r-ranked sets of pairs $\{(A(1), C(1)), \ldots, (A(s), C(s))\}$ with pairwise commuting $n \times n$ $A(\cdot)$ matrices. By Theorem 2, the following hold:

$$\mathcal{N}(s, n, r) \leq N(s, n, r)$$

$$\mathcal{N}_c(s, n, r) \leq N_c(s, n, r).$$

Note that, so far, we have not taken into account p, i.e. the size of the measurements. It should be understood that the numbers $\mathcal{N}_c$ and $\mathcal{N}$ are maxima for all values of p. Likewise, N_c and N are upper bounds for all p. However, throughout the remainder of the chapter, we will assume $p = 1$ for the sake of clarity, but the proofs can easily be modified to account for larger values of p.

Tables I and II contain the upper bounds $N(s, n, n)$ and $N_c(s, n, n)$, respectively, for the first few relevant values of n and s. Note that they grow extremely rapidly (in fact, $N(s, n, n)$ grows so fast that even $N(3, 3, 3)$ is unavailable), which does not really make pathwise observability easy to check. Imagine computing the rank of $s^{N(s,n,n)}$ different $N(s, n, n) \times n$ matrices, even for $s = 2$ and $n = 3$. For now, we simply point out that our upper bounds may be too conservative, and that we leave the

Table 2: Values of $N_c(s,n,n)$ for small n and s

n \ s	1	2	3	4
1	1	1	1	1
2	2	3	4	5
3	3	11	85	1029
4	4	2059	$\simeq 3.610^{40}$	$\simeq 3.310^{619}$

task of reducing them to a future endeavor. In the mean time, note that finding the exact values of $\mathcal{N}(s,n,n)$ and $\mathcal{N}_c(s,n,n)$ is an even more difficult problem, to which the only solution we now have is to match the upper bounds to the actual index of pathwise observability of a particular set of pairs. It is indeed easy to see that $\mathcal{N}(1,n,n) = \mathcal{N}_c(1,n,n) = n$ and that $\mathcal{N}(s,1,1) = \mathcal{N}_c(s,1,1) = 1$. That $\mathcal{N}(s,2,2) = \mathcal{N}_c(s,2,2) = s+1$ follows from the fact that the set of pairs $\{(A,C(1)),\ldots,(A,C(s))\}$, where:

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \ \lambda \neq 1, 0 \\
C(i) &= (1 \ \lambda^{-i}), \ i = 1, \ldots, s,
\end{aligned}
$$

is pathwise observable with index $s+1$. Since $N(s,2,2) = N_c(s,2,2) = s+1$, it follows that $\mathcal{N}(s,2,2) = \mathcal{N}_c(s,2,2) = s+1$.

2.3 Proof of Theorem 2

In this section, we prove Theorem 2. We begin by showing the result for commuting A's (part 2), because its proof is easier and makes use of the main observations in a much more direct way. But before that, we need to establish some preliminary notation.

2.3.1 Preliminaries

In order to make the development more straightforward, we begin with a few definitions. θ^1 and θ^2 being paths of length N_1 and N_2 respectively, $\theta^1\theta^2$ denotes their concatenation $\theta_1^1\theta_2^1\cdots\theta_{N_1}^1\theta_1^2\theta_2^2\cdots\theta_{N_2}^2$. Furthermore, $\theta^{(q)}$ is the path θ concatenated with itself $q-1$ times. Given a path θ, $\theta_{[i,j]}$ is its substring (or infix) $\theta_i\theta_{i+1}\cdots\theta_j$. By convention, we let $\theta_{[i,i-1]} = \epsilon$, the null string, for all $1 \leq i \leq |\theta|$. We also define the transition matrix $\Phi(\theta)$ of a path θ of length N as $\Phi(\theta) = A(\theta_N)\cdots A(\theta_1)$, and note that $\Phi(\theta^1\theta^2) = \Phi(\theta^2)\Phi(\theta^1)$ for any pair of paths θ^1 and θ^2. Again, by convention, we let $\Phi(\epsilon) = I$, the $n \times n$ identity matrix. Finally, let $\mathcal{O}(\theta)_{[i,j]}$ denote the submatrix of $\mathcal{O}(\theta)$ constituted by rows i through j of $\mathcal{O}(\theta)$:

$$\mathcal{O}(\theta)_{[i,j]} \triangleq \begin{pmatrix} C(\theta_i)A(\theta_{i-1})\cdots A(\theta_1) \\ \vdots \\ C(\theta_j)A(\theta_{j-1})\cdots A(\theta_1) \end{pmatrix},$$

and note that

$$\mathcal{O}(\theta)_{[i,j]} = \mathcal{O}(\theta_{[i,j]})\Phi(\theta_{[1,i-1]}), \tag{15}$$

for all i and j such that $1 \leq i \leq j \leq |\theta|$. For the sake of clarity, we will favor the notation of the right hand side of (15). For example, if $\theta = \theta^0\theta^1$, where $|\theta^0| = N_0$ and $|\theta^1| = N_1$, then

$$\mathcal{O}(\theta)_{[N_0+1,N_0+N_1]} = \mathcal{O}(\theta^1)\Phi(\theta^0). \tag{16}$$

Note that the right hand side of (16) is easier to read and makes much more explicit the fact that we are looking at the observability matrix of θ^1 "shifted forward by θ^0." Finally, we let $\mathcal{R}(M)$ denote the row range space of a matrix M.

2.3.2 Proof of Theorem 2, Part 2 (Pairwise Commuting A's)

The fact that $\mathcal{N}_c(1, n, n) = n$ is usually attributed to the Cayley-Hamilton Theorem [40]. However, trying to extend this approach to the switched case has led us nowhere. We therefore need to take another approach, and the following elementary observation actually provides an alternate way to show that $\mathcal{N}_c(1, n, n) = n$. Assume that the rank of the observability matrix does not increase at the kth measurement, i.e. that

$$CA^{k-1} = \sum_{i=1}^{k-1} \alpha_i CA^{i-1}$$

for some $k-1$ real scalars $\{\alpha_i\}_{i=1}^{k-1}$. Right multiplying this equation on both sides by any power of A (i.e. $A^{k'}$), we get

$$CA^{k-1+k'} = \sum_{i=1}^{k-1} \alpha_i CA^{i-1+k'},$$

which implies that the rank has stopped growing for good. $\mathcal{N}(1, n, n) = n$ then follows from the fact that the rank can grow at most n times. It turns out that this argument, along with the Pigeon-Hole Principle alone, is sufficient for establishing the finiteness of $\mathcal{N}_c(s, n, n)$. It should be clear by now that we will prove that pathwise observability is decidable by showing how to construct unobservable paths of arbitrary lengths whenever a system is not pathwise r-ranked at $N_c(s, n, r)$. Our observation translates into the following lemma in the switched case:

Lemma 1 (Range Inclusion Propagation) *Let θ^0, θ^1, and θ^2 be paths of lengths $N_0 \geq 0$, $N_1 > 0$ and $N_2 > 0$ respectively. Assume that:*

$$\mathcal{R}(\mathcal{O}(\theta^2)\Phi(\theta^0\theta^1)) \subset \mathcal{R}(\mathcal{O}(\theta^1)\Phi(\theta^0)). \tag{17}$$

We then have

$$\mathcal{R}(\mathcal{O}(\theta^2)\Phi(\theta^0\theta^3\theta^1)) \subset \mathcal{R}(\mathcal{O}(\theta^1)\Phi(\theta^0\theta^3)) \tag{18}$$

for any path θ^3 of length $N_3 \geq 0$. $\Diamond$

Note that the range inclusion (17) holds between two submatrices of $\mathcal{O}(\theta)$, where $\theta = \theta^0\theta^1\theta^2$, and that (18) concerns $\theta' = \theta^0\theta^3\theta^1\theta^2$. In both cases, the submatrices involved are supported by θ^1 and θ^2, but the difference lies in the fact that $\theta^1\theta^2$ is shifted in θ' by a path θ^3. In other words, what Lemma 1 really tells us is that

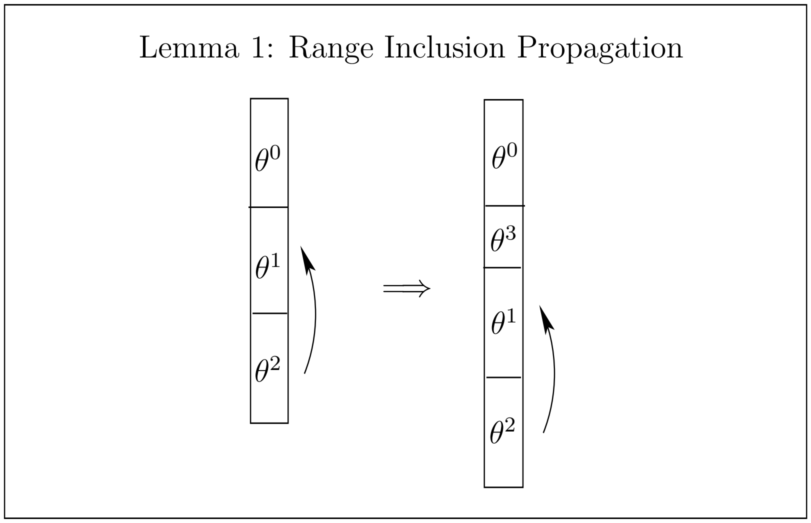


Figure 1: Illustration of Lemma 1. The arrow between two matrix blocks denotes range inclusion

range inclusions within paths are conserved when the paths involved are both equally shifted. An illustration is given in Figure 2.3.2.

The proof is as follows:

Proof: Equation (17) implies the existence, for all $1 \leq k \leq N_2$, of N_1 scalars $\{\alpha_i\}_{i=1}^{N_1}$ such that:

$$C(\theta_k^2)\Phi(\theta_{[1,k-1]}^2)\Phi(\theta^0\theta^1) = \sum_{i=1}^{N_1} \alpha_i C(\theta_i^1)\Phi(\theta_{[1,i-1]}^1)\Phi(\theta^0).$$

Now, by commutativity of the $A(\cdot)$'s and therefore of the $\Phi(\cdot)$'s, and by recalling that $\Phi(\lambda^1\lambda^2) = \Phi(\lambda^2)\Phi(\lambda^1)$ for any pair of paths $\{\lambda^1, \lambda^2\}$, we get for all k, $1 \leq k \leq N_2$,

$$\begin{aligned} C(\theta_k^2)\Phi(\theta_{[1,k-1]}^2)\Phi(\theta^0\theta^3\theta^1) &= C(\theta_k^2)\Phi(\theta_{[1,k-1]}^2)\Phi(\theta^0\theta^1)\Phi(\theta^3) \\ &= \sum_{i=1}^{N_1} \alpha_i C(\theta_i^1)\Phi(\theta_{[1,i-1]}^1)\Phi(\theta^0)\Phi(\theta^3) \\ &= \sum_{i=1}^{N_1} \alpha_i C(\theta_i^1)\Phi(\theta_{[1,i-1]}^1)\Phi(\theta^0\theta^3), \end{aligned}$$

hence (18). $\square$

The following lemma shows how to construct paths of bounded rank of arbitrary length:

Lemma 2 *Let λ^0, λ^1 and λ^2 be paths of lengths $N_0 \geq 0$, $N_1 > 0$ and $N_2 > 0$ respectively. Assume that there exists a path λ^3 such that $\lambda^1\lambda^2 = \lambda^3\lambda^1$, and assume that*

$$r = \rho(\mathcal{O}(\lambda^1)\Phi(\lambda^0)) = \rho(\mathcal{O}(\lambda^1\lambda^2)\Phi(\lambda^0)). \tag{19}$$

Then for any integer m, letting $\lambda' = \lambda^1\lambda^{2^{(m)}}$, we get

$$\mathcal{R}(\mathcal{O}(\lambda')\Phi(\lambda^0)) \subset \mathcal{R}(\mathcal{O}(\lambda^1)\Phi(\lambda^0)), \tag{20}$$

which implies that $\rho(\mathcal{O}(\lambda')\Phi(\lambda^0)) = r$. $\Diamond$

Proof: First, note that for any integer k,

$$\lambda^1\lambda^{2^{(k)}} = \lambda^{3^{(k)}}\lambda^1,$$

which can be shown by induction. We next realize that Equation (19) implies that $\mathcal{R}(\mathcal{O}(\lambda^2)\Phi(\lambda^0\lambda^1)) \subset \mathcal{R}(\mathcal{O}(\lambda^1)\Phi(\lambda^0))$. Therefore, by Lemma 1, we have that

$$\mathcal{R}(\mathcal{O}(\lambda^2)\Phi(\lambda^0\lambda^1\lambda^{2^{(k)}})) = \mathcal{R}(\mathcal{O}(\lambda^2)\Phi(\lambda^0\lambda^{3^{(k)}}\lambda^1)) \subset \mathcal{R}(\mathcal{O}(\lambda^1)\Phi(\lambda^0\lambda^{3^{(k)}})) \tag{21}$$

for $0 \leq k < m$ (simply let $\theta^0 = \lambda^1$, $\theta^1 = \lambda^1$, $\theta^2 = \lambda^2$ and $\theta^3 = \lambda^{3^{(k)}}$). Now, since $\mathcal{O}(\lambda^1)\Phi(\lambda^0\lambda^{3^{(k)}})$ is a submatrix of $\mathcal{O}(\lambda^1\lambda^{2^{(k)}})\Phi(\lambda^0)$, (21) yields

$$\mathcal{R}(\mathcal{O}(\lambda^2)\Phi(\lambda^0\lambda^1\lambda^{2^{(k)}})) \subset \mathcal{R}(\mathcal{O}(\lambda^1\lambda^{2^{(k)}})\Phi(\lambda^0)), \tag{22}$$

which, since $\mathcal{O}(\lambda^1\lambda^{2^{(k+1)}})\Phi(\lambda^0)$ contains both arguments of the *range* function in (22) as submatrices, implies that

$$\mathcal{R}(\mathcal{O}(\lambda^1\lambda^{2^{(k+1)}})\Phi(\lambda^0)) \subset \mathcal{R}(\mathcal{O}(\lambda^1\lambda^{2^{(k)}})\Phi(\lambda^0)), \tag{23}$$

for $0 \leq k < m$. Finally, (23) yields (20) by induction on k, $0 \leq k < m$, and by transitivity of the range inclusion partial ordering. $\square$

We now establish the main lemma of this section:

Lemma 3 *Let θ^0 and θ^1 be two paths of lengths $N_0 \geq 0$ and $N_c(s,n,r)$ respectively. If*

$$\rho(\mathcal{O}(\theta^1)\Phi(\theta^0)) < r,$$

then there exist paths θ^2 of arbitrary lengths N_2 such that

$$\mathcal{R}(\mathcal{O}(\theta^2)\Phi(\theta^0)) \subset \mathcal{R}(\mathcal{O}(\theta^1)\Phi(\theta^0)),$$

resulting in $\rho(\mathcal{O}(\theta^2)\Phi(\theta^0)) < r$. $\Diamond$

Proof: The proof is by induction on r, for $r \leq n$.

Assume that $\theta^1 = t$, $t \in \{1, \ldots, s\}$, and that $C(t)\Phi(\theta^0) = 0$. Let $\theta^2 = t^{(N_2)}$. Then $\mathcal{O}(\theta^2)\Phi(\theta^0) = 0$, because $C(t)\Phi(\theta^0\theta^2_{[1,k]}) = C(t)\Phi(\theta^2_{[1,k]})\Phi(\theta^0) = C(t)\Phi(\theta^0)\Phi(\theta^2_{[1,k]}) = 0$ for all $k \leq N_2$.

Now assume that Lemma 3 is true at $r-1$. We then have two cases:

First, assume there exists $i \in \{0, \ldots, s^{N_c(s,n,r-1)}\}$ such that $\rho(\mathcal{O}(\theta^1)_{[i+1,i+N_c(s,n,r-1)]}\Phi(\theta^0)) < r-1$. Defining $\lambda^0 = \theta^0\theta^1_{[1,i]}$ and $\lambda^1 = \theta^1_{[i+1,i+N_c(s,n,r-1)]}$, Lemma 3 at $r-1$ gives a path λ^2 of arbitrary length such that

$$\mathcal{R}(\mathcal{O}(\lambda^2)\Phi(\lambda^0)) \subset \mathcal{R}(\mathcal{O}(\lambda^1)\Phi(\lambda^0)).$$

Appending the matrix $\mathcal{O}(\theta^1_{[1,i]})\Phi(\theta^0)$ on top of both $\mathcal{O}(\lambda^2)\Phi(\lambda^0)$ and $\mathcal{O}(\lambda^1)\Phi(\lambda^0)$, and noting that since $\theta^1_{[1,i]}\lambda^1 = \theta^1_{[1,i+N_c(s,n,r-1)]}$, $\mathcal{O}(\theta^1_{[1,i]}\lambda^1)\Phi(\theta^0)$ is a submatrix of $\mathcal{O}(\theta^1)\Phi(\theta^0)$, we finally get

$$\mathcal{R}(\mathcal{O}(\theta^1_{[1,i]}\lambda^2)\Phi(\theta^0)) \subset \mathcal{R}(\mathcal{O}(\theta^1)\Phi(\theta^0)),$$

which concludes this case, since λ^2 is of arbitrary length.

Second, assume for all $i \in \{0, \ldots, s^{N_c(s,n,r-1)}\}$, $\rho(\mathcal{O}(\theta^1)_{[i+1,i+N_c(s,n,r-1)]}\Phi(\theta^0)) = r-1$, which implies that $\rho(\mathcal{O}(\theta^1)\Phi(\theta^0)) = r-1$. Furthermore, since there are $s^{N_c(s,n,r-1)}$ different paths of length $N_c(s,n,r-1)$, and since the cardinality of $\{0, \ldots, s^{N_c(s,n,r-1)}\}$

is $s^{N_c(s,n,r-1)} + 1$, there exist, by virtue of the Pigeon Hole Principle, $i, j \in \{0, \ldots, s^{N_c(s,n,r-1)}\}$, $i < j$, such that:

$$\theta^1_{[i+1, i+N_c(s,n,r-1)]} = \theta^1_{[j+1, j+N_c(s,n,r-1)]}.$$

Letting $\lambda^0 = \theta^0 \theta^1_{[1,i]}$, $\lambda^1 = \theta^1_{[i+1, i+N_c(s,n,r-1)]}$, $\lambda^2 = \theta^1_{[i+N_c(s,n,r-1)+1, j+N_c(s,n,r-1)]}$, and $\lambda^3 = \theta^1_{[i+1,j]}$, we have $\lambda^1\lambda^2 = \lambda^3\lambda^1$. Moreover, (19) holds since, by assumption, the range of $\mathcal{O}(\theta^1)_{[i+1,i+N_c(s,n,r-1)]}\Phi(\theta^0)$ spans that of $\mathcal{O}(\theta^1)\Phi(\theta^0)$. Lemma 2 thus gives us a path λ' of arbitrary length such that

$$\mathcal{R}(\mathcal{O}(\lambda')\Phi(\lambda^0)) \subset \mathcal{R}(\mathcal{O}(\lambda^1)\Phi(\lambda^0)).$$

By the same argument as in Case 1, we have

$$\mathcal{R}(\mathcal{O}(\theta^1_{[1,i]}\lambda')\Phi(\theta^0)) \subset \mathcal{R}(\mathcal{O}(\theta^1)\Phi(\theta^0)),$$

which completes the proof. □

An illustration of Lemma 3 is given in Figure 2.3.2.

We can now prove part 2 of Theorem 2:

Proof: Assume that there exists a path θ^1 of length $N_c(s, n, r)$, but that $\rho(\mathcal{O}(\theta^1)) < r$. Assuming $\theta^0 = \epsilon$, Lemma 3 directly implies the existence of paths of arbitrary length of rank strictly smaller than r. □

2.3.3 Proof of Theorem 2, Part 1 (The General Case)

We now address the general case, where we assume nothing about the $A(\cdot)$ matrices. The loss of commutativity destroys the previous results, even though it can be shown that the same upper bounds hold when the A matrices are all invertible, but not necessarily pairwise commuting. Nevertheless, Lemmas 1 and 2 easily carry over to the general case slightly modified, yielding the following two weaker lemmas whose proofs we omit to conserve space.

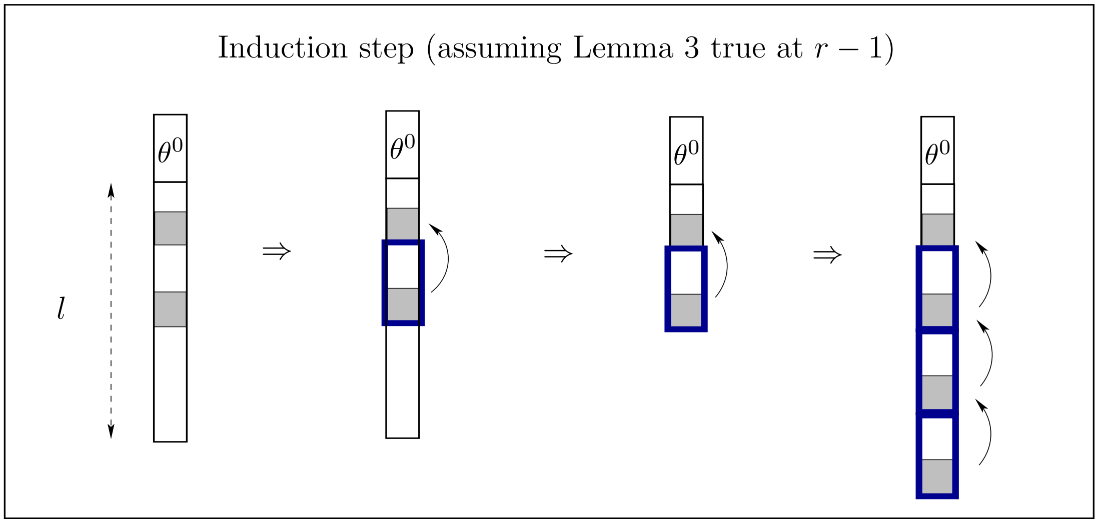


Figure 2: Illustration of the inductive step in Lemma 3. The length l of θ^1 satisfies $l = N_c(s,n,r)$. The grey path is λ^1, and is repeated by the Pigeon-Hole Principle. The path enclosed in blue is the path λ^2 that can be repeated indefinitely to yield a path of arbitrary length not increasing the rank of the observability matrix, by Lemma 2.

Lemma 4 *Let θ^1 and θ^2 be paths of lengths $N_1 > 0$ and $N_2 > 0$ respectively. Assume that:*

$$\mathcal{R}(\mathcal{O}(\theta^2)\Phi(\theta^1)) \subset \mathcal{R}(\mathcal{O}(\theta^1)).$$

We then have

$$\mathcal{R}(\mathcal{O}(\theta^2)\Phi(\theta^3\theta^1)) \subset \mathcal{R}(\mathcal{O}(\theta^1)\Phi(\theta^3))$$

for any path θ^3 of length $N_3 \geq 0$. $\Diamond$

Lemma 5 *Let λ^1 and λ^2 be paths of lengths $N_1 > 0$ and $N_2 > 0$ respectively, assume that there exists a path λ^3 such that $\lambda^1\lambda^2 = \lambda^3\lambda^1$, and assume that*

$$r = \rho(\mathcal{O}(\lambda^1)) = \rho(\mathcal{O}(\lambda^1\lambda^2)).$$

Then for any integer m, letting $\lambda' = \lambda^1\lambda^{2^{(m)}}$, we get

$$\mathcal{R}(\mathcal{O}(\lambda')) \subset \mathcal{R}(\mathcal{O}(\lambda^1)),$$

which implies that $\rho(\mathcal{O}(\lambda')) = r$. $\Diamond$

We now unfortunately need a few more definitions. We say that a path θ is generated by the set $\{\lambda_1, \ldots, \lambda_p\}$ of paths of length l if for any $k \in \{0, \ldots, N - l\}$, where $N = |\theta|$, there exists some $i \in \{1, \ldots, p\}$ such that $\theta_{[N-k-l+1, N-k]} = \lambda_i$. The language generated by the set $\{\lambda_1, \ldots, \lambda_p\}$ is the language containing all paths generated by $\{\lambda_1, \ldots, \lambda_p\}$. We now define:

Definition 4 (Conditional Pathwise r-rank) *The set of pairs $\{(A(1), C(1)), \ldots, (A(s), C(s))\}$ is L-conditionally pathwise r-ranked if and only if there exists an integer N such that all paths of length N in the language L are of rank r. The smallest such integer N_r is called the index of L-conditional pathwise r-rank.* $\Diamond$

We now fix s and n, and define:

Definition 5 *$\mathcal{G}(r, g, p, k)$ is the maximum index of L-conditional pathwise k-rank, over all languages L generated by p paths of length g, and such that g is larger than or equal to the index of L-conditional $r-1$-rank of L.* $\Diamond$

We are now in measure to show the following lemma:

Lemma 6 *$\mathcal{G}(r, g, p, k) \leq G(r, g, p, k)$ for $k = r, \ldots, n$, for $p = 1, \ldots, s^g$, and for all values of r and g required to compute $\mathcal{N}(s, n, r)$.* $\Diamond$

Proof: The proof is by a double induction over k and p, and it suffices to show the following:

(i) $\mathcal{G}(r, g, 1, r) \leq G(r, g, 1, r)$.

(ii) If

$$\mathcal{G}(r, g, p, k) \leq G(r, g, p, k), \quad \text{and} \tag{24}$$

$$\mathcal{G}(k+1, G(r, g, p, k), s^{G(r,g,p,k)}, k+1) \leq G(k+1, G(r, g, p, k), s^{G(r,g,p,k)}, k+1), \tag{25}$$

$$\text{then} \quad \mathcal{G}(r, g, p, k+1) \leq G(r, g, p, k+1). \tag{26}$$

(iii) If $\mathcal{G}(r,g,p,k) \leq G(r,g,p,k)$, $k = r,\ldots,n$, then

$$\mathcal{G}(r,g,p+1,k) \leq 1 + \max_{k=r,r+1,\ldots,n} \{G(r,g,p,k)\}.$$

(i) Let λ be a path of length g such that $\rho(\mathcal{O}(\lambda)) = r-1$. First, λ can generate a path θ of length larger than g if and only if λ is the constant path (i.e. $\lambda = t^{(k)}$ for some $t \in \{1,\ldots,s\}$ and some integer k). Let then θ be a path of length $g+1$ generated by λ such that $\rho(\mathcal{O}(\theta)) = r-1$. Then any path $\theta' = t^{(m)}$ of arbitrary length satisfies $\rho(O(\theta')) = r-1$ by Lemma 5.

(ii) It is clear, from the definition of $\mathcal{G}$, that

$$\mathcal{G}(r,g,p,k+1) \leq \mathcal{G}(k+1, \mathcal{G}(r,g,p,k), s^{\mathcal{G}(r,g,p,k)}, k+1).$$

Given that $\mathcal{G}(r,g,s^g,k)$ is nondecreasing in g and assuming (24), and then assuming (25) is true, we get

$$\mathcal{G}(r,g,p,k+1) \leq G(k+1, G(r,g,p,k), s^{G(r,g,p,k)}, k+1),$$

which yields (26) by definition of $G(r,g,p,k+1)$.

(iii) All we need to show is that

$$\mathcal{G}(r,g,p+1,r) \leq 1 + \max_{k=r,\ldots,n} \{\mathcal{G}(r,g,p,k)\},$$

because the conclusion follows from $\mathcal{G}(r,g,p,k) \leq G(r,g,p,k)$ and from $G(r,g,p+1,r) = 1 + \max_{k=r,\ldots,n}\{G(r,g,p,k)\}$ (by definition of G). Let θ be a path of length $N = 1 + \max_{k=r,\ldots,n}\{\mathcal{G}(r,g,p,k)\}$. Assume that it is generated by $p+1$ paths of length g, and that g is greater than the index of L-conditional $r-1$-rank of L, the language generated by those $p+1$ paths. Assume also that $\rho(\mathcal{O}(\theta)) < r-1$. We then have two cases. First, assume that $\rho(\mathcal{O}(\theta_{[1,g]})) < r-1$. Then, by definition of g, the system would not even be pathwise $r-1$-ranked. Second, assume that $\rho(\mathcal{O}(\theta)) = \rho(\mathcal{O}(\theta_{[1,g]})) = r-1$. We make the first remark that if $\theta_{[1,g]}$ appears elsewhere in θ, say $\theta_{[q+1,q+g]} = \theta_{[1,g]}$ for some q, then, by Lemma 5, $\theta_{[1,g]}\theta^{(m)}_{[g+1,g+q]}$ has rank $r-1$ for all m, and thus the system is not pathwise r-ranked. We can therefore assume that θ contains only one occurrence of $\theta_{[1,g]}$, and we note that if $\rho(A(\theta_1)) = r_0$, then, given that $\mathcal{O}(\theta)_{[2,N]} = A(\theta_1)\mathcal{O}(\theta_{[2,N]})$,

$$\rho(\mathcal{O}(\theta_{[2,N]})) \geq \rho(\mathcal{O}(\theta)_{[2,N]}) \geq \rho(\mathcal{O}(\theta_{[2,N]})) - (n - r_0). \tag{27}$$

Let $n_0 = r + (n - r_0)$. Now, if $\rho(\mathcal{O}(\theta)) < r$, then $\rho(\mathcal{O}(\theta)_{[2,N]}) < r$. If $n_0 \leq n$, then (27) gives $\rho(\mathcal{O}(\theta_{[2,N]})) < n_0$, which, since $\theta_{[2,N]}$ is generated by only p paths of length g and $|\theta_{[2,N]}| \geq \mathcal{G}(r, g, p, n_0)$, implies the existence of a path θ' of arbitrary length such that $\mathcal{R}(\mathcal{O}(\theta')) \subset \mathcal{R}(\mathcal{O}(\theta_{[2,N]}))$, and thus $\rho(\mathcal{O}(\theta_1\theta')) < r$. If $n_0 > n$, then either $\rho(\mathcal{O}(\theta_{[2,N]})) < n$ (see previous case) or $\mathcal{R}(A(\theta_1)) \subset \mathcal{R}(\mathcal{O}(\theta))$, or in other words $A(\theta_1)$ annihilates any chances of increasing $\rho(\mathcal{O}(\theta))$. □

And finally, the proof of Theorem 2, part 1:

Proof: The proof is by induction on r, for $r \leq n$.

Clearly, $\mathcal{N}(s, n, 1) = 1 \leq N(s, n, 1) = 1$.

Now, assuming $\mathcal{N}(s, n, r-1) \leq N(s, n, r-1)$, we get

$$\begin{aligned} \mathcal{N}(s, n, r) &= \mathcal{G}(r, \mathcal{N}(s, n, r-1), s^{\mathcal{N}(s,n,r-1)}, r) && (28)\\ &\leq G(r, \mathcal{N}(s, n, r-1), s^{\mathcal{N}(s,n,r-1)}, r) && (29)\\ &\leq G(r, N(s, n, r-1), s^{N(s,n,r-1)}, r), && (30) \end{aligned}$$

hence $\mathcal{N}(s, n, r) \leq N(s, n, r)$ by definition of $N(s, n, r)$. Equation (28) follows from the definition of $\mathcal{N}$ and $\mathcal{G}$, (29) from Lemma 6, and (30) from the fact that $G(r, g, s^g, r)$ is nondecreasing in g and $\mathcal{N}(s, n, r-1) \leq N(s, n, r-1)$. □

2.4 Pathwise Controllability

Let us recall the model:

$$x_{k+1} = A(\theta_k)x_k + B(\theta_k)u_k, \ k \geq 1$$

whose controllability properties can be characterized by the set of pairs $\{(A(1), B(1)), \ldots, (A(s), B(s))\}$. Defining the controllability matrix of a path θ of length N as

$$\mathcal{C}(\theta) \triangleq (B(\theta_N) \ \ldots \ A(\theta_N) \cdots A(\theta_2) B(\theta_1)),$$

we note that

$$\mathcal{C}(\theta)' \triangleq \begin{pmatrix} B(\theta_N)' \\ B(\theta_{N-1})' A(\theta_N)' \\ \vdots \\ B(\theta_1)' A(\theta_2)' \cdots A(\theta_N)' \end{pmatrix}$$

happens to be equal to the observability matrix of the reversed path θ', where $\theta'_i = \theta_{N-i+1}$, obtained with the set of dual pairs $\{(A(1)', B(1)'), \ldots, (A(s)', B(s)')\}$. By defining *pathwise controllability* as pathwise observability of the set of dual pairs, all our previous results thus carry over to pathwise controllability, and we get:

Theorem 3 *Assume given a set of s pairs $\{(A(1), B(1)), \ldots, (A(s), B(s))\}$, where the dimension of the $A(\cdot)$ matrices is n.*

1. *If the set is pathwise r-ranked, then its index of pathwise r-rank is smaller than or equal to $N(s, n, r)$.*

2. *If furthermore the $A(\cdot)$ matrices are pairwise commuting, then the index of pathwise r-rank is bounded by the smaller number $N_c(s, n, r)$.* $\Diamond$

Corollary 2 *The pathwise r-rank and pathwise controllability properties are decidable.* $\Diamond$

2.5 Conclusion

In this chapter, we have shown that pathwise observability and controllability are decidable. Unfortunately, the upper bounds given are too large to be of any practical significance, and it remains unknown whether they are actually reached.

CHAPTER 3

PATHWISE OBSERVABILITY: FURTHER RESULTS

3.1 Introduction

In this chapter, we restrict our attention o the following subclass of SLS's:

$$
\begin{aligned}
x_{k+1} &= A x_k \\
y_k &= C(\theta_k) x_k,
\end{aligned}
\tag{31}
$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^p$, and where the matrices A and $C(\cdot)$ are of compatible dimensions. The modes θ_k assume values in $\{1, \ldots, s\}$, so that $C(\theta_k)$ switches among s different measurement matrices $C(1), \ldots, C(s)$. The system in (31) can be used for modeling switches between s different sensory modes, as can occur, e.g., when sensors fail intermittently, or when the measurements y_k are transmitted over a memoryless erasure channel [29, 62]. In [62], estimators were designed for the noisy counterpart of (31), and in [9], an asymptotic observer was proposed. In this chapter, we are concerned with a particular aspect of the deterministic finite-time observability of the model, namely *pathwise observability*, whose definition we recall next.

Let a path θ of length N be a string of length N, whose elements take values in $\{1, \ldots, s\}$, and let $|\theta| = N$ denote its length. Defining the observability matrix $\mathcal{O}(\theta)$ of a path θ as:

$$
\mathcal{O}(\theta) \triangleq \begin{pmatrix} C(\theta_1) \\ \vdots \\ C(\theta_N) A^{N-1} \end{pmatrix},
\tag{32}
$$

we say that θ is observable when its observability matrix is of full rank. If we let

$\rho(M)$ denote the rank of a matrix M, this condition can thus be written as

$$\rho(\mathcal{O}(\theta)) = n. \tag{33}$$

We now arrive at the definition of *pathwise observability*:

Definition 6 (Pathwise Observability) *The set of pairs* $\{(A, C(1)), \ldots, (A, C(s))\}$ *is pathwise observable if and only if there exists an integer* N *such that all paths of length* N *are observable. We refer to the smallest such integer as the index of pathwise observability.* $\diamondsuit$

We showed in the previous chapter that pathwise observability is decidable. In fact, since A commutes with itself, we showed that the indexes of pathwise observability are bounded by numbers $\mathcal{N}_c(s, n, n)$ depending only on s and n, which is an even stronger result, since it suggests a direct way of deciding whether or not a set of pairs is pathwise observable, in checking the rank of the observability matrix of every path of length $\mathcal{N}_c(s, n, n)$. In this chapter, we will give sufficient conditions for pathwise observability that allow us to come up with a switched version of the Kalman-Bertram criterion for non-pathological sampling.

In Section 3.2, we establish sufficient conditions based on structural properties of the individual pairs, which is an interesting result in that it dispenses from computing coupled observability matrices (32) (i.e. matrices involving multiple modes), enabling the study of classical observability matrices of standard dimensions. In Section 3.3, we use that result to extend a classical result from linear systems theory concerning the conservation of observability properties when sampling a continuous-time system. In Section 3.4, we dualize these results to the controllability case.

3.2 Sufficient Conditions for Pathwise Observability

In this section, we establish sufficient conditions on the individual pairs $(A, C(i))$ for the set $\{(A, C(1)), \ldots, (A, C(s))\}$ to be pathwise observable. More precisely, the idea is that if a pair $(A^b, C(i))$ is observable, $b \in \mathbb{N}$, then whenever $\theta_{a+bk} = i$ for $k = 0, \ldots, n-1$ and for some integer a, i.e. whenever some mode i occurs n times in θ at constant interval b, then $\mathcal{O}(\theta)$ will contain the following matrix as a submatrix:

$$\begin{pmatrix} C(i) \\ C(i)A^b \\ \vdots \\ C(i)(A^b)^{n-1} \end{pmatrix} A^{a-1}, \tag{34}$$

which has rank n if A is invertible, and therefore ensures that $\rho(\mathcal{O}(\theta)) = n$. Note that this would not be the case if there were switching among different A-matrices as well. In that case, the matrix in (34) would, in general, still exhibit coupling with modes other than i. What we thus want to show is that whenever a pair $(A^l, C(i))$ is observable for all modes i and for all l smaller than a certain number, then the system is pathwise observable. This implies the possibility to assert that, in every path of at least a certain length $\mathcal{W}$, some mode i has to occur n equally separated times. It turns out that proving the existence of such $\mathcal{W}$ is a problem to which an answer is provided by a branch of combinatorial analysis, referred to as *Ramsey theory* [35]. Indeed, we wish to capitalize on the fact that any mode sequence has to exhibit certain regularity properties as long as it is long enough, which is a type of statement that falls precisely under the domain of Ramsey theory, whose main assertion is that complete disorder is an impossibility and that the appearance of disorder is really a matter of scale. As it turns out, our question finds its answer in van der Waerden's Theorem [69] (in its finite version), which is one of the central results of Ramsey theory:

Theorem 4 (van der Waerden [69]) *For every positive integers n and s, there exists a minimal constant $\mathcal{W}(n, s)$ such that if $N \geq \mathcal{W}(n, s)$, and $\{1, \ldots, N\} \subset C_1 \cup \ldots \cup C_s$, then some set C_i contains an arithmetic progression of length n.* $\Diamond$

Here, an arithmetic progression is simply a string of positive integers such that the difference between successive terms is constant. It is indeed easy to see how the solution to our problem follows from Theorem 4 by simply taking every C_i to be the set of times at which mode i occurs in θ. In other words, if we ignore the trivial case $n = 1$ and assume $n \geq 2$, which will be done throughout the remainder of this chapter, we have:

Corollary 3 *Let θ be a path over $\{1, \ldots, s\}$. If $|\theta| \geq \mathcal{W}(n, s)$, then there exist an integer $i \in \{1, \ldots, s\}$ and two positive integers $a \in \{1, \ldots, |\theta|\}$ and $b < |\theta|/(n-1)$ such that $\theta_{a+bk} = i$ for every $k = 0, \ldots, n-1$.* $\Diamond$

Proof: Let $C_i = \{k \in \{1, \ldots, |\theta|\} \mid \theta_k = i\}$ for all $i \in \{1, \ldots, s\}$. Clearly, $\{1, \ldots, |\theta|\} \subset C_1 \cup \ldots \cup C_s$. By Theorem 4, since $|\theta| \geq \mathcal{W}(n, s)$, some C_i contains an arithmetic progression of length n. In other words, there exist two positive integers a and b such that $a + bk \in C_i$, and therefore $\theta_{a+bk} = i$, for $k = 0, \ldots, n-1$. Finally, $b < |\theta|/(n-1)$ because $b(n-1) < a + b(n-1) \leq |\theta|$. $\square$

An illustration of Corollary 3 is provided in Figure 3.2.

Before establishing the main result of this section, which is a direct consequence of Corollary 3, we define, for $n \geq 2$,

$$\mathcal{W}'(n, s) \triangleq \left\lceil \frac{\mathcal{W}(n, s)}{n-1} \right\rceil - 1, \tag{35}$$

where $\lceil \cdot \rceil$ denotes the ceiling function (i.e. $\lceil \alpha \rceil = \min\{i \in \mathbb{N} \mid \alpha \leq i\}$).

Theorem 5 *If A is invertible, and if $(A^l, C(i))$ is an observable pair for all $i \in \{1, \ldots, s\}$ and all positive integers $l \leq \mathcal{W}'(n, s)$, then $\{(A, C(1)), \ldots, (A, C(s))\}$ is pathwise observable with an index no larger than $\mathcal{W}(n, s)$.* $\Diamond$

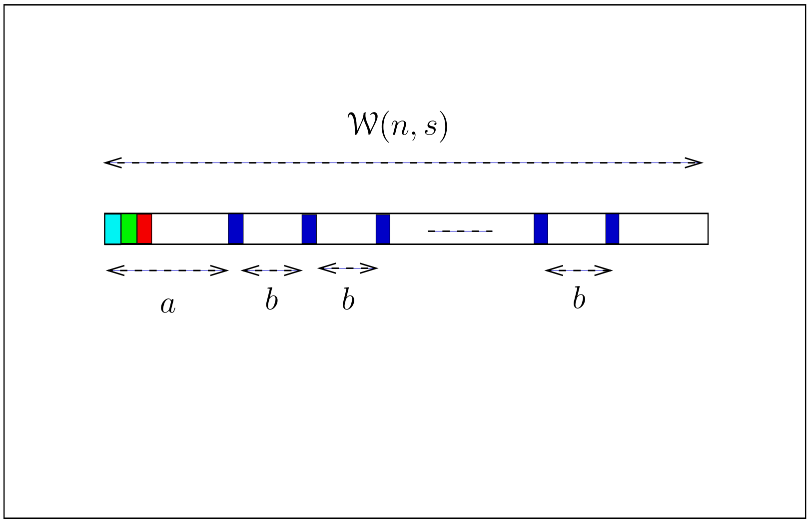


Figure 3: Illustration of Corollary 3. Every color represents a different mode. If the length of the path is greater than $\mathcal{W}(n,s)$, then a mode (here the blue one) is repeated in the path n times at constant interval.

Proof: Let θ be any path of length $\mathcal{W}(n,s)$. By Corollary 3, there exist an integer $i \in \{1,\dots,s\}$ and two integers $a \in \{1,\dots,|\theta|\}$ and $b < \mathcal{W}(n,s)\backslash(n-1)$ such that $\theta_{a+bk} = i$ for $k = 0,\dots,n-1$. Therefore, the submatrix of $\mathcal{O}(\theta)$ consisting of the rows $a+bk$ of $\mathcal{O}(\theta)$, $k = 0,\dots,n-1$, can be expressed as

$$\begin{pmatrix} C(i) \\ C(i)A^b \\ \vdots \\ C(i)(A^b)^{n-1} \end{pmatrix} A^{a-1}.$$

This matrix has rank n since A (and therefore A^{a-1}) is invertible, and because the pair $(A^b, C(i))$ is observable, since $b \le \mathcal{W}'(n,s)$. Therefore $\mathcal{O}(\theta)$ has rank n, which completes the proof. $\square$

Remarks 2

- *These conditions are* not *necessary. For instance, the set of pairs* $\{(A,C(1)),(A,C(2))\}$, *where:*

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{cases} C(1) = (1\ 0) \\ C(2) = (2\ 0) \end{cases} \tag{36}$$

is pathwise observable with index 2*, but while* $\mathcal{W}'(2,2) = 2$*, neither* $(A^2, C(1))$ *nor* $(A^2, C(2))$ *is an observable pair.*

- *The index of pathwise observability in Theorem 5 is not necessarily equal to* $\mathcal{W}(n,s)$*. For instance, the set of pairs* $\{(A, C(1)), (A, C(2))\}$*, where:*

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{cases} C(1) = (1\ 1) \\ C(2) = (1\ 2) \end{cases} \tag{37}$$

satisfies the assumptions of Theorem 5, but is pathwise observable with index 2*, while* $\mathcal{W}(2,2) = 3$*.* $\Diamond$

The numbers $\mathcal{W}(n,s)$ are referred to as the van der Waerden (vdW) numbers. Unfortunately, the only vdW numbers known exactly fit in Table I (for the sake of easy reference, we also give, in Table II, the known values of $\mathcal{W}'(n,s)$). Only upper

Table 3: Known values of $\mathcal{W}(n,s)$

s \ n	2	3	4	5	$\cdots$	n
1	2	3	4	5	$\cdots$	n
2	3	9	35	178		
3	4	27				
4	5	76				
$\vdots$	$\vdots$					
s	$s+1$					

Table 4: Known values of $\mathcal{W}'(n,s)$

s \ n	2	3	4	5	$\cdots$	n
1	1	1	1	1	$\cdots$	1
2	2	4	11	88		
3	3	13				
4	4	37				
$\vdots$	$\vdots$					
s	s					

bounds are known for the rest. Those bounds grow at an enormous rate, which limits the applicability of Theorem 5. In fact, research is currently ongoing for finding tighter bounds, e.g. [34, 60]. However, Theorem 5 is fortunately all we need in order to show the more practical results of the next section concerning sampled systems.

3.3 Sampled-Data Systems

A problem of relevance to digital control is the study of properties of sampled-data systems since most modern, digital controllers are implemented in discrete-time. In particular, it is usually desirable for a discretized system to conserve some properties of the continuous-time system, especially observability and controllability. We start, without loss of generality, by considering the following autonomous, continuous-time system:

$$\begin{aligned} \dot{x}_t &= Ax_t \\ y_t &= Cx_t, \end{aligned} \tag{38}$$

and the discrete-time system obtained by sampling (38) at constant interval T, which is referred to as the sampling period (for any continuous-time quantity z_t, we let $\bar{z}_k \triangleq z_{kT}$):

$$\begin{aligned} \bar{x}_{k+1} &= e^{AT}\bar{x}_k \\ \bar{y}_k &= C\bar{x}_k. \end{aligned} \tag{39}$$

In 1963, the following result was proved in [41]:

Theorem 6 (Kalman-Bertram Criterion) *Let $\sigma(A)$ denote the spectrum of A. If (A, C) is an observable pair, then whenever the sampling period T satisfies, for all $\{\lambda, \lambda'\} \in \sigma(A) \times \sigma(A)$,*

$$\lambda \neq \lambda' + \frac{jk}{T} \quad \forall\ k \in \mathbb{Z}\backslash\{0\}, \tag{40}$$

then the discrete-time pair (e^{AT}, C) is observable. $\Diamond$

A proof can be found in [41], but the result easily follows from the Popov-Belevitch-Hautus rank test (see, e.g., [22]). Further research on this subject has focused mainly on generalized hold functions [39, 57] (for controllability) and on robust sampling techniques [46].

Our aim in this section is to extend Theorem 6 to switched linear systems. In other words, we focus our attention on the continuous-time switched linear system:

$$\begin{aligned} \dot{x}_t &= Ax_t \\ y_t &= C(\theta_t)x_t, \end{aligned} \tag{41}$$

where θ_t is an arbitrary function of time assuming values in the set $\{1, \ldots, s\}$, and on its discretized counterpart:

$$\begin{aligned} \bar{x}_{k+1} &= e^{AT}\bar{x}_k \\ \bar{y}_k &= C(\bar{\theta}_k)\bar{x}_k. \end{aligned} \tag{42}$$

Note that, even though θ_t is arbitrary and may switch between samples, (42) can be characterized by a finite set of pairs $\{(e^{AT}, C(1)), \ldots, (e^{AT}, C(s))\}$, which cannot be the case when the dynamics (i.e. the A matrix) switches as well (unless, e.g., θ_t switches only at the sampling times). What we wish to establish here is whether observability of every pair $(A, C(i))$ implies pathwise observability of the set of pairs $\{(e^{AT}, C(1)), \ldots, (e^{AT}, C(s))\}$. Fortunately, the following theorem follows almost directly from Theorems 5 and 6:

Theorem 7 *Let $\sigma(A)$ denote the spectrum of A. If $(A, C(i))$ is an observable pair for all $i \in \{1, \ldots, s\}$, then whenever the sampling period T satisfies, for all $\{\lambda, \lambda'\} \in \sigma(A) \times \sigma(A)$,*

$$\lambda \neq \lambda' + \frac{jk}{lT} \quad \forall\ k \in \mathbb{Z}\backslash\{0\},\ \forall\ l \leq \mathcal{W}'(n, s), \tag{43}$$

the set of pairs $\{(e^{AT}, C(1)), \ldots, (e^{AT}, C(s))\}$ of the discretized system is pathwise observable with an index no larger than $\mathcal{W}(n, s)$. $\Diamond$

Proof: First, since AT commutes with itself and l is an integer, $e^{AlT} = (e^{AT})^l$. Therefore, by Theorem 6, (43) implies that the pair $((e^{AT})^l, C(i))$ is observable for all $i \in \{1,\ldots,s\}$ and all $l \leq \mathcal{W}'(n,s)$. Moreover, e^{AT} being a matrix exponential, it is an invertible matrix. The result then follows from Theorem 5. $\square$

Now, even though some numbers $\mathcal{W}(n,s)$ may be unknown, they are finite, as discussed earlier. The following corollary follows:

Corollary 4 *If $(A, C(i))$ is an observable pair for all $i \in \{1,\ldots,s\}$, then the set of pairs $\{(e^{AT}, C(1)),\ldots,(e^{AT}, C(s))\}$ of the discretized system is pathwise observable for all but a countable number of sampling periods T.* $\Diamond$

Proof: If every eigenvalue of A is real, then (43) always holds and the set is pathwise observable for all $T > 0$. Otherwise, defining the set F of frequencies as

$$F \triangleq \{|\mathrm{Im}(\lambda_i) - \mathrm{Im}(\lambda_j)| \;|\lambda_i \neq \lambda_l \in \sigma(A),\ \mathrm{Re}(\lambda_i) = \mathrm{Re}(\lambda_j)\}, \tag{44}$$

we get that the set of pathological sampling periods is, by (43), a subset of

$$\left\{\frac{k}{fl},\ k \in \mathbb{N}^*,\ f \in F,\ l \leq \mathcal{W}'(n,s)\right\}, \tag{45}$$

which is countable. Hence the result. $\square$

Finally, note that what needs to be avoided in Theorem 6 is the interaction between the natural frequencies of the linear system and the sampling frequency. It is therefore easily established that, under the same conditions, conservation of observability is guaranteed when the sampling period T is small enough. The importance of this observation is actually further motivated by robust control problems, as pointed out in [46]. The following theorem extends this result to switched linear systems (41):

Theorem 8 *If $(A, C(i))$ is an observable pair for all $i \in \{1,\ldots,s\}$, then there exists a positive real number T such that whenever $0 < t < T$, the set of pairs $\{(e^{At}, C(1)),\ldots,(e^{At}, C(s))\}$ of the discretized system is pathwise observable with an index smaller than or equal to $\mathcal{W}(n,s)$.* $\Diamond$

Proof: Clearly,

$$T = \frac{1}{\max(F)\mathcal{W}'(n,s)}, \tag{46}$$

which is the smallest element of the set in (45), works. $\square$

The most surprising fact about Theorems 7 and 8 is that there is inherently no coupling between the s different modes in continuous-time, and yet pathwise observability is shown to be achieved for the sampled-data system. Moreover, note that we make absolutely no assumption on θ_t, other than that it is a mapping from the continuous time line to $\{1,\ldots,s\}$. In particular, T in Theorem 8 is an upper bound on the sampling period, and *not* a lower bound on the switching intervals (or minimum *dwell time*).

3.4 Pathwise Controllability

Notice that the first results of this chapter naturally carry over, by duality, to the study of switched systems of the form:

$$x_{k+1} = Ax_k + B(\theta_k)u_k, \tag{47}$$

where the modes θ_k assume values in $\{1,\ldots,s\}$, so that $B(\theta_k)$ switches among s different input matrices $\{B(1),\ldots,B(s)\}$, and where one may be concerned with *pathwise controllability*, defined as pathwise observability of the set of dual pairs $\{((A',B(1)'),\ldots,(A',B(s)')\}$ [5]. In fact, one gets, as a trivial extension of Theorem 5:

Theorem 9 *If A is invertible, and if $(A^l, B(i))$ is a controllable pair for all $i \in \{1,\ldots,s\}$ and all integers $l \leq \mathcal{W}'(n,s)$, then $\{(A,B(1)),\ldots,(A,B(s))\}$ is pathwise controllable with an index no larger than $\mathcal{W}(n,s)$.* $\lozenge$

However, one should be careful when considering the sampling problem from the controllability point of view. Indeed, applying a *zero-order hold* to

$$\dot{x}_t = Ax_t + B(\theta_t)u_t, \tag{48}$$

i.e. letting $u_t \triangleq \bar{u}_k \ \forall\ t \in [kT, (k+1)T)$, yields

$$\bar{x}_{k+1} = e^{AT}\bar{x}_k + B_k\bar{u}_k, \tag{49}$$

where $B_k = \int_{kT}^{(k+1)T} e^{A((k+1)T-\tau)}B(\theta_\tau)d\tau$. Once again, B_k might switch among an infinite number of values, unless, e.g., the signal θ_t is constrained to switch at only the sampling times. In fact, the dual of our criterion (Theorem 7) involves the use of a Dirac impulse-based discretization as follows:

$$u_t = \bar{u}_k\delta(t - kT), \qquad kT \leq t < (k+1)T, \tag{50}$$

which allows us to rewrite (49) as

$$\bar{x}_{k+1} = e^{AT}\bar{x}_k + B(\theta_k)\bar{u}_k, \tag{51}$$

to which we can then apply the previous results. Now, even though (50) does not make any sense since perfect impulses cannot be produced in practice, we can state the following purely theoretical result:

Theorem 10 *Let $\sigma(A)$ denote the spectrum of A. If $(A, B(i))$ is a controllable pair for all $i \in \{1, \ldots, s\}$, then whenever the sampling period T satisfies, for all $\{\lambda, \lambda'\} \in \sigma(A) \times \sigma(A)$,*

$$\lambda \neq \lambda' + \frac{jk}{lT} \quad \forall\ k \in \mathbb{Z}\backslash\{0\},\ \forall\ l \leq \mathcal{W}'(n, s), \tag{52}$$

the set of pairs $\{(e^{AT}, B(1)), \ldots, (e^{AT}, B(s))\}$ of the discretized system (51) obtained by applying the hold function (50) to (48) is pathwise controllable with an index no larger than $\mathcal{W}(n, s)$. ◊

Finally, note that Corollary 4 also extends to the controllability case.

3.5 Conclusions

In this chapter, we have introduced an application of Ramsey Theory to the study of a property of switched linear systems (i.e. *pathwise observability*). The result presented has enabled, for the first time, the study of the conservation of observability and controllability properties after the introduction of sampling in switched systems, which has resulted in a criterion very similar to the well-known Kalman-Bertram criterion.

CHAPTER 4

OBSERVABILITY UNDER UNKNOWN MODES

4.1 Introduction

By switched linear systems (SLS), we refer to discrete-time systems that can be modeled as follows:

$$\begin{aligned} x_{k+1} &= A(\theta_k)x_k + B(\theta_k)u_k \\ y_k &= C(\theta_k)x_k, \end{aligned} \tag{53}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, and $y_k \in \mathbb{R}^p$ are the states, the inputs and the measurements, respectively. θ_k, which we refer to as the mode at time k, assumes its values in the set $\{1, \ldots, s\}$, so that the system parameter matrices $A(\theta_k)$, $B(\theta_k)$, and $C(\theta_k)$ switch among s different known matrices. We assume that the mode sequence $\{\theta_k\}_{k=1}^{\infty}$, whether known or unknown, is arbitrary and independent of the initial state and inputs. In particular, we impose no constraints on the time separation between two consecutive switches, and we assume that the switches are not triggered by state space based events.

By observability, we mean the ability to infer the initial state x_1, and possibly a finite portion of the mode sequence (when unobserved), from a finite number of measurements $y_1, \ldots, y_N$. While the concept of observability has a simple well-known characterization in classical linear systems, it has been associated with several notions in the SLS literature. Indeed, the fact that the mode sequence may or may not be observed, and, in the latter case, that one may or may not wish to recover it along with the state, makes for the need to consider several different problems, thus different definitions and characterizations. In this chapter, our aim is to introduce and to define several different concepts of observability in SLS's under unknown modes, to characterize them, and to assess their main properties, among which decidability is of special importance.

The outline of this chapter is as follows. In Section 4.2, the autonomous case is studied. In Section 4.3, we discuss the non-autonomous case.

4.2 Autonomous Systems

We restrict our attention, for now, to autonomous systems of the form:

$$\begin{aligned} x_{k+1} &= A(\theta_k)x_k \\ y_k &= C(\theta_k)x_k, \end{aligned} \tag{54}$$

obtained simply by removing the $B(\theta_k)u_k$ term from (53). Before going any further, we need a few definitions. Since we are dealing with finite-time problems, we abandon the mode sequence notation and we define a path θ as a finite sequence (or string) of modes $\theta = \theta_1\theta_2 \dots \theta_N$, where N is the path length denoted by $|\theta|$. We also define Θ_N as the set of all paths of length N. Moreover, we denote by $\theta_{[i,j]}$ the infix of θ between i and j, i.e. $\theta_{[i,j]} = \theta_i\theta_{i+1} \dots \theta_j$, we use $\theta\theta'$ to denote the concatenation of θ with θ', and we let $\Phi(\theta) \triangleq A(\theta_N) \cdots A(\theta_1)$ denote the transition matrix of a path θ. By convention, we let $\theta_{[i,i-1]} = \epsilon$, the empty word, and $\Phi(\epsilon) = I$. We next define the observability matrix $\mathcal{O}(\theta)$ of a path θ as:

$$\mathcal{O}(\theta) \triangleq \begin{pmatrix} C(\theta_1) \\ \vdots \\ C(\theta_N)A(\theta_{N-1}) \cdots A(\theta_1) \end{pmatrix}. \tag{55}$$

Finally, we define:

$$Y(\theta, x) \triangleq \mathcal{O}(\theta)x, \tag{56}$$

and we then get that if $x = x_1$ and $\theta = \theta_1\theta_2 \dots \theta_N$ in (54), then $Y(\theta, x) = (y_1^T \ \dots \ y_N^T)^T$. Therefore, throughout the remainder of this section, we will use (56) to describe (54) in a more compact way.

In this section, we will define and characterize several concepts of observability for autonomous SLS (54). We first present, in Subsection 4.2.1, a preliminary result that will later be used in the decidability proofs. In Subsections 4.2.2 and 4.2.3, we study mode observability and state observability, respectively.

4.2.1 A Preliminary Result

In this subsection, we let

$$N(s,n) \triangleq N(s,n,n), \tag{57}$$

i.e. the upper bound on $\mathbb{N}(s,n,n)$ given in 2.2, and we state the following corollary to Theorem 2, whose proof can easily be derived from that of Theorem 2:

Theorem 11 *If θ is a path of length $N(s,n)$, then there exists a prefix θ^0 of θ (i.e. $\theta = \theta^0\theta^1$ for some θ^1) and a path θ' of arbitrary length such that*

$$\mathcal{R}(\mathcal{O}(\theta^0\theta')) \subset \mathcal{R}(\mathcal{O}(\theta^0)), \tag{58}$$

and thus $\rho(\mathcal{O}(\theta^0\theta')) = \rho(\mathcal{O}(\theta^0)) \leq \rho(\mathcal{O}(\theta))$. $\diamondsuit$

4.2.2 Mode Observability

In this section, we assume that only the continuous measurements $Y(\theta,x)$ are available, and we investigate the possibility to infer a prefix of the path θ (i.e. $\theta_{[1,N']}$ for some $N' < |\theta|$) from the successive measurements $Y(\theta,x)$ only. But first, noting that when $x = 0$, $Y(\theta,x) = 0$ for any path θ, we observe that it is *impossible* to distinguish between paths whenever $x = 0$. As it turns out, this happens in general for all states in a union of subspaces of $\mathbb{R}^n$. Moreover, this issue is closely related to false alarms in failure detection, as pointed out in [11]. We therefore have to consider the problem from a looser point of view, which leads us to the following definition, in which *a.e.* x stands for "for almost every x", by which we mean for all $x \in \mathbb{R}^n$ but a union of proper subspaces, thus for all x but a set of Lebesgue measure 0:

Definition 7 (Mode Observability (MO)) *The SLS (54) is MO at N if there exists an integer N' such that for all $\theta \in \Theta_{N+N'}$ and for a.e. $x \in \mathbb{R}^n$,*

$$\theta_{[1,N]} \neq \theta'_{[1,N]} \Rightarrow Y(\theta, x) \neq Y(\theta', x') \;\forall\; x' \in \mathbb{R}^n \tag{59}$$

The index of MO at N is the smallest such N'. $\diamondsuit$

In other words, we require the possibility to recover the first N modes (i.e. $\theta_{[1,N]}$) uniquely whenever $N+N'$ measurements (i.e. $Y(\theta, x)$) are available, and for a.e. state x. To this end, we need a way to discern between the paths θ using the measurements $Y(\theta, x)$ they produce through $Y(\theta, x) = \mathcal{O}(\theta)x$. As we are about to show, the only way to achieve that without any information other than the available measurement $Y(\theta, x)$ is by taking advantage of the following inclusion, immediate from $Y(\theta, x) = \mathcal{O}(\theta)x$:

$$Y(\theta, x) \in \Re(\mathcal{O}(\theta)), \tag{60}$$

where $\Re(M)$ denotes the column range space of the matrix M. The question is then whether $\theta' \neq \theta \;\Rightarrow\; Y(\theta, x) \notin \Re(\mathcal{O}(\theta'))$, which would provide us with a simple procedure for recovering a path from the measurements (using the range inclusion test, see Appendix A):

$$\theta = \arg_{\theta' \in \Theta_N} \{Y(\theta, x) \in \Re(\mathcal{O}(\theta'))\} \tag{61}$$

The main issue lies in whether the test (61) has a unique solution. In order to analyze this, we introduce the concept of *discernibility*:

Definition 8 (Discernibility) *A path θ is discernible from another path θ' of the same length if*

$$\rho([\mathcal{O}(\theta)\mathcal{O}(\theta')]) > \rho(\mathcal{O}(\theta')), \tag{62}$$

where $[\mathcal{O}(\theta)\mathcal{O}(\theta')]$ denotes the horizontal concatenation of $\mathcal{O}(\theta)$ and $\mathcal{O}(\theta')$, and where the degree d of discernibility is defined as

$$d = \rho([\mathcal{O}(\theta)\mathcal{O}(\theta')]) - \rho(\mathcal{O}(\theta')). \tag{63}$$

We then say that θ is d-discernible from θ'. $\Diamond$

The following proposition is now in order:

Proposition 1 *$Y(\theta, x) \notin \Re(\mathcal{O}(\theta'))$ for almost any $x \in \mathbb{R}^n$ iff θ is discernible from θ'.* $\Diamond$

Proof: We have $Y(\theta, x) \in \Re(\mathcal{O}(\theta'))$ iff $Y(\theta, x)$ also lies in the the *output subspace of conflict of θ and θ'*, defined as:

$$C(\theta, \theta') \triangleq \Re(\mathcal{O}(\theta)) \cap \Re(\mathcal{O}(\theta')). \tag{64}$$

We therefore need to show that the dimension of the inverse image of $C(\theta, \theta')$ by $\mathcal{O}(\theta)$, $c(\theta, \theta') \triangleq \mathcal{O}(\theta)^{-1}(C(\theta, \theta'))$, which we refer to as the *input subspace of conflict of θ with θ'*, is smaller than n (which implies that its Lebesgue measure is 0) if and only if θ is discernible from θ'. We have:

$$\dim(C(\theta, \theta')) = \rho(\mathcal{O}(\theta)) + \rho(\mathcal{O}(\theta')) - \rho([\mathcal{O}(\theta)\mathcal{O}(\theta')]). \tag{65}$$

Noting that $\dim(c(\theta, \theta')) = \dim(C(\theta, \theta')) + \dim\ker(\mathcal{O}(\theta))$, and then recalling that $\rho(\mathcal{O}(\theta)) + \dim\ker(\mathcal{O}(\theta)) = n$, we get

$$\dim(c(\theta, \theta')) = n - \rho([\mathcal{O}(\theta)\mathcal{O}(\theta')]) + \rho(\mathcal{O}(\theta')). \tag{66}$$

Therefore, by definition of discernibility, we see that $\dim(c(\theta, \theta')) < n$ if and only if θ is discernible from θ', in which case we moreover have

$$\dim(c(\theta, \theta')) = n - d, \tag{67}$$

where d is the degree of discernibility defined in (63). $\square$

Remarks 3

- *Discernibility does not imply that either path is observable (see Example 1).*

- *From (67), the degree of discernibility appears as a measure of* separation *between the two paths θ and θ': the larger the degree d, the smaller the dimension of the input subspace of conflict. It is clear that the maximum value for d is n, in which case the input subspace of conflict is trivial.*

- *Note that, as we have defined it, discernibility is not symmetric.*

- *If $\theta_{[1,N-1]} = \theta'_{[1,N-1]}$, i.e. if the two paths only differ by their last value, then their index of discernibility is bounded by p.* $\Diamond$

The last remark raises the question of whether an upper bound (p, the size of each measurement y_k) is imposed on the maximum degree of discernibility that can be guaranteed for all pairs of paths of a certain length. It turns out that this limitation can be overcome, provided one can use further measurements in order to discern the paths, which leads us to the idea of *forward discernibility*:

Definition 9 (Forward Discernibility (FD)) *Given an integer $d > 0$, a path θ is forward d-discernible (d-FD) from another path θ' of the same length if there exists an integer N_d such that for any pair of paths λ and λ' of length N_d, $\theta\lambda$ and $\theta'\lambda'$ are discernible with degree at least d. The smallest such integer N_d is the index of d-FD of θ from θ'.* $\Diamond$

Proposition 2 *$Y(\theta\lambda, x) \notin \Re(\mathcal{O}(\theta'\lambda'))$ for all $\lambda, \lambda' \in \Theta_{N'}$ and for almost any $x \in \mathbb{R}^n$ iff θ is FD (i.e. 1-FD) from θ' with an index no larger than N'.* $\Diamond$

Proof: Clearly, the set $\{x \in \mathbb{R}^n \mid \exists \lambda, \lambda' \in \Theta_{N'},\ Y(\theta\lambda, x) \in \Re(\mathcal{O}(\theta'\lambda'))\}$ equals $\bigcup_{\lambda,\lambda'} c(\theta\lambda, \theta'\lambda')$, which, by Proposition 1 and by virtue of the fact that a finite union of null sets is a null set, has measure 0 iff θ is FD from θ'. $\square$

We now turn to showing that forward discernibility is decidable. We first establish the following lemma, which indicates that the indexes of d-FD increase with d, which is not an obvious fact.

Lemma 7 *Let θ and θ' be two different paths of length N, and λ and λ' be any paths of length N'. The degree of discernibility of $\theta\lambda$ from $\theta'\lambda'$ is greater than or equal to the degree of discernibility of θ from θ'. In other words, the degree of discernibility is nondecreasing as the length increases.* $\diamondsuit$

Proof: It is easily shown, by elementary linear algebra, that

$$\rho([\mathcal{O}(\theta\lambda)\mathcal{O}(\theta'\lambda')]) - \rho([\mathcal{O}(\theta)\mathcal{O}(\theta')]) \geq \rho(\mathcal{O}(\theta\lambda)) - \rho(\mathcal{O}(\theta)). \tag{68}$$

In other words, the rank of the concatenation must increase by at least the increase of each path. $\square$

Theorem 12 (Decidability of Forward Discernibility) *FD is decidable for any degree, as the index of d-FD, where d is the maximum degree of FD between some pair of paths, is smaller than or equal to $N(s^2, 2n)$.* $\diamondsuit$

Proof: Fix θ and θ', and let λ and λ' be such that the degree of discernibility of $\theta\lambda$ from $\theta'\lambda'$ is minimal over all pairs of paths λ and λ' of length $N(s^2, 2n)$.

First, note that the matrices $[\mathcal{O}(\lambda)\mathcal{O}(\lambda')]$ are produced by the following set of s^2 pairs:

$$\begin{pmatrix} A(i) & 0 \\ 0 & A(j) \end{pmatrix} \qquad \begin{pmatrix} C(i) & C(j) \end{pmatrix} \qquad (i,j) \in \{1, \ldots, s\}^2. \tag{69}$$

Therefore, by Theorem 11, there exist λ^0 and λ'^0, respective prefixes of λ and λ' of the same length, and two paths μ and μ' of the same, arbitrary, length, such that $\mathcal{R}([\mathcal{O}(\lambda^0\mu)\mathcal{O}(\lambda'^0\mu')]) \subset \mathcal{R}([\mathcal{O}(\lambda^0)\mathcal{O}(\lambda'^0)])$, which, by [5, Lemma 4] and upon some manipulation, implies that

$$\mathcal{R}([\mathcal{O}(\theta\lambda^0\mu)\mathcal{O}(\theta'\lambda'^0\mu')]) \subset \mathcal{R}([\mathcal{O}(\theta\lambda^0)\mathcal{O}(\theta'\lambda'^0)]). \tag{70}$$

By Lemma 7, Equation (70) implies that the degree of discernibility of $\theta\lambda^0\mu$ from $\theta'\lambda'^0\mu'$ is equal to that of $\theta\lambda^0$ from $\theta'\lambda'^0$, which, again by Lemma 7, is smaller than that of $\theta\lambda$ from $\theta'\lambda'$, which completes the proof since μ and μ' are of arbitrary length. $\square$

Before establishing the main result of this section characterizing mode observability, we need the following definition:

Definition 10 (Complete Forward Discernibility (CFD)) *Given an integer $d > 0$, a path θ is completely forward d-discernible (d-CFD) if it is d-FD from every other path θ' of the same length. The index of d-CFD of θ is the maximum index of d-FD, over all $\theta' \neq \theta$, of θ from θ'.* $\diamondsuit$

Theorem 13 *The SLS (54) is MO at N iff every path of length N is CFD, (i.e. 1-CFD). Moreover, the index of MO is the largest index of CFD, over all paths of length N.* $\diamondsuit$

Proof: Clearly, the set $\{x \in \mathbb{R}^n \mid \exists\theta,\theta' \in \Theta_N,\ \exists\lambda,\lambda' \in \Theta_{N'},\ Y(\theta\lambda, x) \in \Re(\mathcal{O}(\theta'\lambda'))\}$ equals $\bigcup_{\theta,\theta'}\bigcup_{\lambda,\lambda'} c(\theta\lambda,\theta'\lambda')$. Therefore, by Proposition 2, and by virtue of the fact that a finite union of null sets is a null set, it has measure 0 iff every θ is CFD with index at most N'. $\square$

We now complete our study of mode observability in autonomous systems by answering the following two questions:

- what effect does N have on MO? In other words, is MO at larger N stronger or weaker?
- Is MO decidable?

The following proposition answers the first question:

Proposition 3 *If a system is MO at N, then it is MO at any $M \leq N$.* $\diamondsuit$

Proof: Let N' be the index of MO at N. Then for every pair of paths θ, θ' of length $N + N'$ with a switch at or before N (i.e. such that $\theta_i \neq \theta'_i$ for some $i \leq N$), θ must be discernible from θ'. But, since $M \leq N$, this implies the same whenever a switch occurs at or before M, which implies MO at M with an index smaller than or equal to $N' + (N - M)$. □

The converse is unfortunately not true, unless the A matrices are all invertible (a counterexample is given next):

Proposition 4 *If $A(1), \ldots, A(s)$ are all invertible, then MO at 1 implies MO at any positive integer N.* ◊

Proof: Let θ and θ' be two different paths of length N, and assume that the maximum index of FD over all pairs of different modes (i.e. paths of length 1) is N'. It suffices to show that θ is FD from θ' with index at most N'. Let λ and λ' be any two paths of length N'. It is easy to show that

$$c(\theta\lambda, \theta'\lambda') = \bigcap_{i=1}^{N} \phi(\theta_{[1,i-1]})^{-1} c(\theta_{[i,N]}\lambda, \theta'_{[i,N]}\lambda'). \tag{71}$$

Since $\theta \neq \theta'$, there exists $i \leq n$ such that $\theta_i \neq \theta'_i$. Note that $c(\theta_{[i,N]}\lambda, \theta'_{[i,N]}\lambda') = c(\theta_i\mu, \theta'_i\mu')$ with $\mu = \theta_{[i+1,N]}\lambda$, $\mu' = \theta'_{[i+1,N]}\lambda'$ and $|\mu| = |\mu'| \geq N'$. Therefore, since, by assumption, θ_i is FD from θ'_i with index at most N', $\dim(c(\theta_{[i,N]}\lambda, \theta'_{[i,N]}\lambda')) < n$, and since all the A matrices, and thus $\phi(\theta_{[1,i-1]})$, are invertible, and using (71), we finally get $\dim(c(\theta\lambda, \theta'\lambda')) < n$, which completes the proof. □

Example 1 *Consider*

$$\begin{aligned} A(1) &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad & A(2) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ C(1) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \quad & C(2) &= \begin{pmatrix} 1 & 0 \end{pmatrix}. \end{aligned} \tag{72}$$

The paths 1 and 2 (of length 1) are mutually FD with index 1, but the paths $1 \cdot 1$ and $1 \cdot 2$ are not, because they are not discernible and $A(1) = 0$, which prevents further measurements from increasing their discernibility. △

And finally,

Theorem 14 *MO at any index N is decidable.* $\Diamond$

Proof: Since the number of paths of length N is finite, and since, by Theorem 12, FD is decidable, it follows that CFD is decidable, and thus that MO is decidable as well. $\square$

Note that more precise versions of these last 3 results can be obtained, provided one extends the notion of *degree* to MO.

4.2.3 State Observability

In this section, we are concerned with whether the continuous state x only is recoverable. From the previous results, we know that if a system is PWO (with index N_{pwo}) and MO at N_{pwo} with index N_{mo}, then one can recover the state x uniquely from $Y(\theta, x)$ for all θ of length $N_{pwo} + N_{mo}$ and for almost any x. But if we do not need θ (which is the primary reason behind the "a.e."), is this still the best we can do? It turns out that we can do better, in that we can sometimes recover *all* states x uniquely, for all paths θ of a certain length. For now, we define our concept of *state observability*:

Definition 11 (State Observability (SO)) *The SLS (54) is SO if there exists an integer N (the smallest being the index) such that $\forall x \in \mathbb{R}^n$ and $\forall \theta \in \Theta_N$,*

$$x \neq x' \Rightarrow Y(\theta, x) \neq Y(\theta', x') \ \forall \ \theta' \in \Theta_N \tag{73}$$

In other words, a system is SO if any N consecutive measurements $Y(\theta, x)$ yield x uniquely without knowledge of θ, i.e. if the map $(x, \theta) \mapsto Y(\theta, x)$ is injective in its first coordinate. We first establish a sufficient condition.

Proposition 5 *If a system is PWO with index N_{pwo}, and if every path of length N_{pwo} is n-CFD, then it is SO.* $\Diamond$

Proof: Let N_{cfd} be the maximum index of n-CFD, and $N = N_{pwo}+N_{cfd}$. This implies that the dimension of the input subspaces of conflict of any two paths of length N satisfying $\theta_{[1,N_{PWO}]} \neq \theta'_{[1,N_{PWO}]}$ is 0: 0 is therefore the only state whose measurements $Y(\theta, x)$ do not yield $\theta_{[1,N_{PWO}]}$ unambiguously. We then have two cases:

- $x \neq 0$, in which case the range inclusion test yields $\theta_{[1,N_{PWO}]}$, which can then be used in $x = \mathcal{O}(\theta_{[1,N_{PWO}]})^{\{1\}}Y(\theta, x)$.
- $x = 0$, in which case $Y(\theta, x) = 0$. By pathwise observability, we then know that $x = 0$. □

Example 2 *As an example, here is a system satisfying the conditions of Proposition 5:*

$$\begin{aligned} A(1) &= \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad A(2) = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix} \\ C(1) &= \begin{pmatrix} 3 & 0 \end{pmatrix} \quad C(2) = \begin{pmatrix} 2 & 0 \end{pmatrix}, \end{aligned} \tag{74}$$

Here, $N_{pwo} = 2$ and $N_{cfd} = 2$, and it is easy to check that the rank of $[\mathcal{O}(\theta\lambda)\mathcal{O}(\theta'\lambda')]$ equals 4 for any pair θ, θ' of different paths of length 2 and any pair λ, λ' of paths of length 2. △

It turns out that the conditions given in Proposition 5 are not necessary (we will later give a counterexample). In order to study SO further, we introduce the concept of *joint observability*:

Definition 12 (Joint Observability (JO)) *Two different paths θ and θ' of the same length are jointly observable (JO) if they are both observable, and if their left inverses agree on $C(\theta, \theta')$, i.e.*[1]

$$(\mathcal{O}(\theta)^{\{1\}} - \mathcal{O}(\theta')^{\{1\}})P_{C(\theta,\theta')} = 0, \tag{75}$$

[1] *Given a subspace V, we let P_V denote the matrix of a linear projection on V.*

or equivalently,

$$(\mathcal{O}(\theta) - \mathcal{O}(\theta'))P_{c(\theta,\theta')} = 0 \text{ and } (\mathcal{O}(\theta) - \mathcal{O}(\theta'))P_{c(\theta',\theta)} = 0. \tag{76}$$

Note that, as opposed to discernibility, joint observability is symmetric. A direct consequence of this definition is:

Proposition 6 *θ and θ' are JO iff for all $x, x' \in \mathbb{R}^n$,*

$$x \neq x' \Rightarrow Y(\theta, x) \neq Y(\theta', x'). \tag{77}$$

We also need to define *forward joint observability*:

Definition 13 (Forward Joint Observability (FJO)) *Two different observable paths θ and θ' of the same length are forward jointly observable (FJO) if there exists an integer N such that for all λ and λ' of length N, $\theta\lambda$ and $\theta'\lambda'$ are JO. The index of FJO is the smallest such integer.* $\Diamond$

Before characterizing SO, we next show that FJO is decidable.

Theorem 15 *FJO is decidable, as the index of JO is bounded by $N(s^2, 2n)$.* $\Diamond$

Proof: Suppose that θ and θ' are observable and that there exist λ and λ' of length $N(s^2, 2n)$ such that $\theta\lambda$ and $\theta'\lambda'$ are not JO. Similarly as in the proof of Theorem 12, we can find λ^0 and λ'^0, respective prefixes of λ and λ' of the same length, and two paths μ and μ' of the same, arbitrary, length, such that

$$\mathcal{R}([\mathcal{O}(\theta\lambda^0\mu)\mathcal{O}(\theta'\lambda'^0\mu')]) \subset \mathcal{R}([\mathcal{O}(\theta\lambda^0)\mathcal{O}(\theta'\lambda'^0)]). \tag{78}$$

Now, since $\theta\lambda$ and $\theta'\lambda'$ are not JO, neither can be $\theta\lambda^0$ and $\theta'\lambda'^0$, since $Y(\theta\lambda, x) = Y(\theta'\lambda', x')$ implies $Y(\theta\lambda^0, x) = Y(\theta'\lambda'^0, x')$.

Moreover, by Lemma 7, Equation (78) implies that the degree of discernibility of $\theta\lambda^0\mu$ from $\theta'\lambda'^0\mu'$ equals that of $\theta\lambda^0$ from $\theta'\lambda'^0$, which furthermore implies that $c(\theta\lambda^0\mu, \theta'\lambda'^0\mu') = c(\theta\lambda^0, \theta'\lambda'^0)$, thus that $(\mathcal{O}(\theta\lambda^0\mu) - \mathcal{O}(\theta'\lambda'^0\mu'))P_{c(\theta\lambda^0\mu,\theta'\lambda'^0\mu')}$ equals

$(\mathcal{O}(\theta\lambda^0\mu)-\mathcal{O}(\theta'\lambda'^0\mu'))P_{c(\theta\lambda^0,\theta'\lambda'^0)}$ and cannot equal zero since its submatrix $(\mathcal{O}(\theta\lambda^0)-\mathcal{O}(\theta'\lambda'^0))P_{c(\theta\lambda^0,\theta'\lambda'^0)}$ is not, because, as we have just shown, $\theta\lambda^0$ and $\theta'\lambda'^0$ are not JO. Therefore, $\theta\lambda^0\mu$ and $\theta'\lambda'^0\mu'$ are not JO, which completes the proof since μ and μ' are of arbitrary length. $\square$

We now characterize SO:

Theorem 16 *The following are equivalent.*

1. *The SLS (54) is SO.*
2. *The SLS (54) is PWO with index N_{pwo}, and every pair of different paths of length N_{pwo} is FJO.*
3. *The SLS (54) is PWO, and every* minimally observable *path (i.e. a path with no observable prefix) is FJO with every other observable path of the same length.* $\Diamond$

Proof:

$2 \Rightarrow 1$: Let N_{fjo} be be the largest index of FJO over all pairs of paths of length N_{pwo}. Let us show that the system is SO with index at most $N = N_{pwo} + N_{fjo}$. Fix a path θ of length N, and suppose that θ' is such that $Y(\theta, x) = Y(\theta', x')$. Let $\theta_{[1,k]}$ be the minimally observable prefix of θ. First, if $\theta'_{[1,k]} = \theta_{[1,k]}$, then $x = x'$ by observability of $\theta_{[1,k]}$, since $Y(\theta, x) = Y(\theta', x')$ implies $Y(\theta_{[1,k]}, x) = Y(\theta'_{[1,k]}, x') = Y(\theta_{[1,k]}, x')$. On the other hand, if $\theta'_{[1,k]} \neq \theta_{[1,k]}$, then since $k \leq N_{pwo}$, it is easy to show that $\theta'_{[1,k]}$ and $\theta_{[1,k]}$ are FJO with index at most $N - k$. Proposition 6 then concludes that $x = x'$.

$3 \Rightarrow 2$: It is easily seen that the only pairs of paths of length N_{pwo} left to check for FJO are those sharing the same minimally observable prefix. Let θ and θ' be two paths of length N_{pwo}, and let $\theta'_{[1,k]} = \theta_{[1,k]}$ be their minimally observable prefix. $Y(\theta, x) = Y(\theta', x')$, which implies $Y(\theta_{[1,k]}, x) = Y(\theta'_{[1,k]}, x') = Y(\theta_{[1,k]}, x')$, implies that $x = x'$ by observability of $\theta_{[1,k]}$. θ and θ' are therefore JO, thus FJO.

$1 \Rightarrow 3$: Necessity of PWO to SO is obvious. Suppose that a minimally observable path is not FJO with another observable path, i.e. that there exist λ, λ' of arbitrary length such that $\theta\lambda$ and $\theta\lambda'$ are not JO, which, by proposition 6, implies the existence of $x \neq x'$ such that $Y(\theta\lambda, x) = Y(\theta'\lambda', x')$, which contradicts SO. $\square$

The reason we give two characterizations is that their equivalence is not obvious, and because the second one is easier to check, since the number of minimally observable paths is in general smaller than $s^{N_{pwo}}$. Moreover, it is, in a sense, much tighter, since two paths can be non FJO only if they do not share the same minimally observable path. Finally,

Theorem 17 *SO is decidable.* $\diamondsuit$

Proof: PWO is decidable. Since FJO is decidable, and since there is a finite number of paths of length N_{pwo}, the first characterization of Theorem 16 concludes. $\square$

We now give an example of an SO system that does not satisfy the requirements of Proposition 5:

Example 3 *Let*

$$\begin{array}{ll} A(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & A(2) = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \\ C(1) = \begin{pmatrix} 1 & 0 \end{pmatrix} & C(2) = \begin{pmatrix} 1 & 0 \end{pmatrix}, \end{array} \tag{79}$$

This system is PWO with index 2, and any paths of length 2 are FJO with index 1. For instance, letting $\theta = 11$, $\theta' = 22$, and $\lambda = \lambda' = 1$, one gets

$$\mathcal{O}(\theta\lambda) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad \mathcal{O}(\theta'\lambda') = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 8 \end{pmatrix}, \tag{80}$$

hence that $\theta\lambda$ and $\theta'\lambda'$ are JO because the first columns of their observability matrices, which span $C(\theta, \theta')$, are equal. Thus if we measure $Y(\mu, x) = (\alpha\ \alpha\ \alpha)^T$, then the initial state can only be $(\alpha\ 0)^T$, regardless of the path μ. $\triangle$

4.3 Non-Autonomous Systems

We now return to the general non-autonomous case, and recall our model:

$$
\begin{aligned}
x_{k+1} &= A(\theta_k)x_k + B(\theta_k)u_k \\
y_k &= C(\theta_k)x_k.
\end{aligned}
\tag{81}
$$

Our aim here is to extend some of the previous analysis to the system in (81). We thus first define, for a path θ of length N:

$$
\mathcal{G}(\theta) \triangleq \begin{pmatrix}
0 & \cdots & 0 & 0 \\
C(\theta_2)B(\theta_1) & \cdots & 0 & 0 \\
C(\theta_3)A(\theta_2)B(\theta_1) & \cdots & \vdots & 0 \\
\vdots & \cdots & 0 & \vdots \\
C(\theta_N)\Phi(\theta_{[2,N]})B(\theta_1) & \cdots & C(\theta_N)B(\theta_{N-1}) & 0
\end{pmatrix},
$$

which enables us to further define:

$$
Y(\theta, x, U) \triangleq \mathcal{O}(\theta)x + \mathcal{G}(\theta)U, \tag{82}
$$

where U is a control vector in $\mathbb{R}^{mN}$. Again, if $x = x_1$, $\theta = \theta_1 \cdots \theta_N$, and $U = (u_1^T \ \ldots \ u_N^T)^T$ in (81), then $Y(\theta, x, U) = (y_1^T \ \ldots \ y_N^T)^T$, and we can concentrate on equation (82). In this section, we will only take a first look at mode observability.

Given θ and θ', our objective in the autonomous case has been, roughly speaking, to make the intersection $C(\theta, \theta')$ of $\Re(\mathcal{O}(\theta))$ with $\Re(\mathcal{O}(\theta'))$ as small as possible. Here, Equation (82) suggests that we should rather consider the intersection of the affine subspaces $\Re(\mathcal{O}(\theta)) + \mathcal{G}(\theta)U$ and $\Re(\mathcal{O}(\theta')) + \mathcal{G}(\theta')U$ ($V + v$, where V is a subspace and v a vector of $\mathbb{R}^n$, denotes the affine subspace $\{x + v \mid x \in V\}$), and study what effect U has on it. Recalling the following classic theorem,

Theorem 18 *The intersection of $V + v$ and $V' + v'$ is either empty or equal to $V \cap V' + w$ for some w, in which case it has the dimension of $V \cap V'$.* $\Diamond$

we realize that, while the $\mathcal{G}(\theta)U$ terms cannot increase the degree of discernibility, they can achieve something impossible in the non-autonomous case: they can render the output affine subspaces of θ and θ', i.e. $\Re(\mathcal{O}(\theta)) + \mathcal{G}(\theta)U$ and $\Re(\mathcal{O}(\theta')) + \mathcal{G}(\theta')U$, totally disjoint, which motivates the following definition:

Definition 14 (Strong Mode Observability (SMO)) *The SLS (81) is strongly mode observable (SMO) at N if there exists an integer N' and a vector U such that for all $x \in \mathbb{R}^n$ and all $\theta \in \Theta_{N+N'}$,*

$$\theta_{[1,N]} \neq \theta'_{[1,N]} \Rightarrow Y(\theta, x, U) \neq Y(\theta', x', U) \ \forall \ x' \in \mathbb{R}^n \tag{83}$$

We refer to such a vector U as a discerning control. $\Diamond$

Note that the difference lies in the replacement of "a.e. x" with "$\forall$ x", which is a stronger statement. In order to characterize SMO, we unfortunately need a few more definitions:

Definition 15 (Controlled-Discernibility (CD)) *Two different paths θ and θ' of length N are controlled-discernible (CD) if*

$$(I - P)\big(\mathcal{G}(\theta) - \mathcal{G}(\theta')\big) \neq 0, \tag{84}$$

where P is the matrix of any projection on $\Re([O(\theta)\ \mathcal{O}(\theta')])$.

It can be verified that CD is well-defined, even though P is not unique. However, to fix the ideas, we let $P(\theta, \theta')$ be the matrix of the orthogonal projection on $\Re([O(\theta)\ \mathcal{O}(\theta')])$, throughout the remainder of this section. Furthermore, note that CD is also symmetric. We can now establish the following:

Proposition 7 *If θ and θ' are CD, then there exists a vector U such that*

$$\forall x \in \mathbb{R}^n,\ Y(\theta, x, U) \notin \Re(\mathcal{O}(\theta')) + \mathcal{G}(\theta')U. \tag{85}$$

Even though $\Re(\mathcal{O}(\theta')) + \mathcal{G}(\theta')U$ is an affine subspace, we can still use the range inclusion test, by testing whether $Y(\theta, x, U)) - \mathcal{G}(\theta')U$ is in $\Re(\mathcal{O}(\theta'))$. The proof of Proposition 7 is as follows:

Proof: Let U satisfy $(I - P(\theta, \theta'))\big(\mathcal{G}(\theta) - \mathcal{G}(\theta')\big)U \neq 0$. Then, by elementary linear algebra, $\Re(\mathcal{O}(\theta)) + \mathcal{G}(\theta)U$ and $\Re(\mathcal{O}(\theta')) + \mathcal{G}(\theta')U$ are totally disjoint as affine subspaces of $\mathbb{R}^{pN}$, which completes the proof, since $Y(\theta, x, U) \in \Re(\mathcal{O}(\theta)) + \mathcal{G}(\theta)U$. $\square$

Finally, we define:

Definition 16 (Forward Controlled-Discernibility (FCD)) *Two different paths θ and θ' of length N are forward controlled-discernible (FCD) if there exists an integer N' such that $\theta\lambda$ and $\theta'\lambda'$ are controlled discernible for any pair of paths λ and λ' of length N'. The smallest such integer is the index of FCD.* $\diamondsuit$

Unfortunately, we do not know whether or not FCD is decidable. This is in part due to the fact that, as opposed to $\mathcal{O}(\theta)$, we know little about the structure of $\mathcal{G}(\theta)$. Nevertheless, we can characterize SMO as follows:

Theorem 19 *The SLS (81) is SMO at N iff any two different paths θ and θ' of length N are FCD.* $\diamondsuit$

Proof: Suppose the system is SMO at N with index N'. It follows that there exists a control vector U such that for all $\theta, \theta' \in \Theta_N$, $\theta \neq \theta'$, and $\lambda, \lambda' \in \Theta_{N'}$, $\Re(\mathcal{O}(\theta\lambda)) + \mathcal{G}(\theta\lambda)U$ and $\Re(\mathcal{O}(\theta'\lambda')) + \mathcal{G}(\theta'\lambda')U$ are totally disjoint, which implies that $(I - P(\theta\lambda, \theta'\lambda'))\big(\mathcal{G}(\theta\lambda) - \mathcal{G}(\theta'\lambda')\big)U \neq 0$, thus that $(I - P(\theta\lambda, \theta'\lambda'))\big(\mathcal{G}(\theta\lambda) - \mathcal{G}(\theta'\lambda')\big) \neq 0$, hence FCD of θ and θ' with index at most N'.

Now, let N' be the maximum index of FCD, over all pairs of different paths of length N, and let us show that the system is SMO with index at most N'. We need to show the existence of a vector U in $\mathbb{R}^{m(N+N')}$ such that

$$(I - P(\theta\lambda, \theta'\lambda'))\big(\mathcal{G}(\theta\lambda) - \mathcal{G}(\theta'\lambda')\big)U \neq 0 \tag{86}$$

for all $\theta, \theta' \in \Theta_N$, $\theta \neq \theta'$, and $\lambda, \lambda' \in \Theta_{N'}$. Since every pair θ and θ' of different paths of length N is FCD with index at most N', we get

$$(I - P(\theta\lambda, \theta'\lambda'))\big(\mathcal{G}(\theta\lambda) - \mathcal{G}(\theta'\lambda')\big) \neq 0. \tag{87}$$

Therefore,

$$K = \bigcup_{\theta,\theta',\lambda,\lambda'} \ker\left((I - P(\theta\lambda, \theta'\lambda'))\big(\mathcal{G}(\theta\lambda) - \mathcal{G}(\theta'\lambda')\big)\right) \neq \mathbb{R}^{m(N+N')}, \tag{88}$$

since it is a finite union of proper subspaces of $\mathbb{R}^{m(N+N')}$, by (87). Any control vector $U \in \mathbb{R}^{m(N+N')} \backslash K$ will work in (86), and is therefore discerning. $\square$

It should be noted that the existence of a single discerning control U implies that "almost" any vector of the same length is discerning, as established by (88). Therefore, we have just shown, under Sontag's terminology [63], that:

Theorem 20 *Single experiment mode observability and generic experiment mode observability are equivalent.* $\diamondsuit$

Finally, we describe an SMO system in the next example.

Example 4 *Let*

$$A(1) = A(2) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad C(1) = C(2) = \begin{pmatrix} 1 & 0 \end{pmatrix} \tag{89}$$

$$B(1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad B(1) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \tag{90}$$

Since the observability pairs $(A(1), C(1))$ and $(A(2), C(2))$ are equal, no two paths can be discernible, because all paths of the same length share the exact same observability matrix. However, this system is SMO at $N = 2$, with index $N' = 1$. To see this, it suffices to use Theorem 19 and to establish that $(I - P(\theta\lambda, \theta'\lambda'))\big(\mathcal{G}(\theta\lambda) - \mathcal{G}(\theta'\lambda')\big) \neq 0$ for any two different paths θ and θ' of length 2, and any pair of paths λ and λ' of length 1. $\triangle$

4.4 Conclusion

We have characterized several concepts of observability in switched linear systems through simple linear algebraic tests, and we have shown their decidability in the autonomous case. An assumption underlying all criteria studied was that the mode sequences were arbitrary, which is novel in the sense that most (if not all) previous work assumed constraints on the mode sequences, usually in the form of minimum “dwell times” between switches.

This chapter is intended as an intermediate step towards a better understanding of the observability of switched systems. Indeed, some results need to be refined, some problems still need to be solved, and many extensions are in view. To mention a few, the decidability of *forward controlled-discernibility* (FCD), which seems to be a challenging problem, and the characterization and study of state observability in the non-autonomous case, still need to be addressed. Finally, the investigation of the application of the concept of *discernibility* to asymptotic observer design promises to be fruitful, and we leave it to a future endeavor.

CHAPTER 5

THE DIRECT ALGEBRAIC APPROACH

5.1 Introduction

In this chapter, we consider the following class subclass of SLS's:

$$\begin{aligned} x_{k+1} &= Ax_k \\ y_k &= C(\theta_k)x_k, \end{aligned} \tag{91}$$

where x_k and y_k are in $\mathbb{R}^n$ and $\mathbb{R}$, respectively, where the *mode* θ_k takes on values in $\{1, \ldots, m\}$, and where A, $C(1), \ldots, C(m)$ are constant matrices of compatible dimensions. We further assume that the mode sequence $\{\theta_k\}_{k=0}^{\infty}$ is arbitrary, indexing the measurement equation in such a way that $C(\theta_k)$ switches randomly among $C(1), \ldots, C(m)$, modeling the m different sensory modes. This model has been used to describe sensor failures or lossy transmission channels [19, 29, 62].

We also assume that it is unknown which one of the m different measurement equations is in effect at any given time instant, or in other words, we assume the mode sequence $\{\theta_k\}_{k=0}^{\infty}$ to be unknown, even though a full characterization of the possible measurement matrices $C(1), \ldots, C(m)$ is available. The goal of this work can now be stated as follows: Devise an asymptotic observer for the system (91), where the switching sequence is arbitrary and unknown. It appears that this problem has never been successfully addressed in a systematic manner. When the mode sequence is observed, it is well known (e.g. [47]) that a Kalman filter can, under some conditions, be used as an observer for (91), and recently, an LMI-based approach has been proposed for designing Luenberger-like switching observers [3]. For obvious reasons, these results are not pertinent to this work. However, capitalizing on the latter approach and on failure detection techniques, an observer design methodology was proposed in [11]. Unfortunately, failure detection schemes require the parameter θ to be slowly-varying, which is too restrictive for the problem at hand.

The outline of this chapter is as follows: We introduce the Direct Algebraic Approach (DAA) in Section 5.2, and we construct the DAA-Newton observer in Section 5.3. In Section 5.4, we analyze some geometric aspects of the observer, which later enable us to prove its local exponential convergence in section 5.5. We finally present some numerical results in Section 5.6.

5.2 The Direct Algebraic Approach

In this section, we present the Direct Algebraic Approach (DAA), which was originally proposed in [45], and recently generalized to (91) in [9]. It can be described as follows: since $\theta_k \in \{1, \ldots, m\}$, the measurement equation in (91), $y_k - C(\theta_k)x_k = 0$, implies that

$$\big(y_k - C(1)x_k\big) \cdots \big(y_k - C(m)x_k\big) = 0. \tag{92}$$

Now, defining

$$g_k(x) \triangleq \big(y_k - C(1)x\big) \cdots \big(y_k - C(m)x\big), \tag{93}$$

we have $g_k(x_k) = 0$. We thus propose to shift our attention to designing an observer for the following deterministic system:

$$\begin{aligned} x_{k+1} &= Ax_k \\ g_k(x_k) &= 0, \end{aligned} \tag{94}$$

where $g_k(x_k) = 0$ is the new nonlinear, time-varying, yet deterministic measurement equation given in implicit form. Indeed, g_k is a deterministic polynomial form whose coefficients are determined by the available measurement y_k. Clearly, the uncertainty associated with the randomly switched measurement equation of the original system in (91) has been removed, and the need to determine θ_k circumvented. Note that a similar idea has been successfully applied to the data association problem in multiple target tracking, leading to the so-called SME filter [43]. Also, similar ideas can be found in fault detection and isolation [74, 2], and in system identification for multi-modal systems [70].

Unfortunately, the transformation of (91) into (94) does come with a price. The price one has to pay for the introduction of a nonlinear measurement equation is that local convergence is in general all one can hope for. In the next section, we complete the construction by combining the DAA with a nonlinear observer, thus obtaining an observer for our original system (91).

5.3 The DAA-Newton Observer

In [58] was proposed a nonlinear observer design approach based on Newton's method, which we refer to as the Newton observers approach. As we will see, a Newton observer can successfully be combined with the DAA, and the key idea is to fix an integer $N_B \geq n$, defined as the "block size", and define a new measurement map as follows:

$$G_k(x) \triangleq \begin{pmatrix} g_k(x) \\ \vdots \\ g_{k+N_B-1}(A^{N_B-1}x) \end{pmatrix} \tag{95}$$

$$= \begin{pmatrix} \prod_{i=1}^{m}(y_k - C(i)x) \\ \vdots \\ \prod_{i=1}^{m}(y_{k+N_B-1} - C(i)A^{N_B-1}x) \end{pmatrix}. \tag{96}$$

Since we have

$$G_k(x_k) = 0, \tag{97}$$

Equation (97) can be used as a new measurement equation, replacing $g_k(x_k) = 0$ in (94) as follows:

$$\begin{aligned} x_{k+1} &= Ax_k \\ G_k(x_k) &= 0. \end{aligned} \tag{98}$$

We can now define the DAA-Newton observer for (91) as:

$$\hat{x}_k^- = A\hat{x}_{k-1} \tag{99}$$

$$\hat{x}_k = \hat{x}_k^- - \left(G_k'(\hat{x}_k^-)\right)^\dagger (G_k(\hat{x}_k^-)), \tag{100}$$

where $G_k'(x)$ is the Jacobian of $G_k(x)$, and where $J^\dagger$ is defined for any full-column rank matrix J as

$$J^\dagger \triangleq (J^T J)^{-1} J^T, \tag{101}$$

and coincides with the pseudo-inverse of J. This implies that $G_k'(\hat{x}_k^-)$ must have full column rank, and sufficient conditions for this to be satisfied are given in the next section.

The observer given by (99-100) is a direct interpretation of the Newton observers approach applied to the system in (98): Equation (100), the "corrector" part of the observer, materializes a single iteration of Newton's method on (97) using $\hat{x}_k^-$ as the initial estimate of the root of G_k, exhibiting the "map inversion" viewpoint of [58]. The motivation behind the construction of G_k in (95) thus becomes obvious: Newton's method cannot be shown to converge to x_k if (97) is underdetermined, hence the condition $N_B \geq n$.

Note that, by construction of G_k, future measurements must be available, i.e. the measurements $y_k, \ldots, y_{k+N_B-1}$ must be available for the computation of $\hat{x}_k$. This limitation can easily be overcome by adding a predictor after (100), i.e. by letting $z_k = A^{N_B-1}\hat{x}_{k-N_B+1}$ be the estimate of x_k, which results in a causal observer. For the sake of simplicity, we will study the convergence of $\hat{x}_k$ rather than that of z_k, and we observe that z_k will converge exponentially to x_k whenever $\hat{x}_k$ does, since:

$$\|z_k - x_k\| \leq \|A\|^{N_B-1} \|\hat{x}_{k-N_B+1} - x_{k-N_B+1}\|. \tag{102}$$

5.4 *Observability and Non-Degeneracy*

In this section, we address two questions concerning the DAA-Newton observer:

1. When does the equation $G_k(x) = 0$, the measurement equation in (98), admit a unique solution, at least locally (i.e. isolated root)?

2. When does $G'_k(x)$ have full column rank, so that its left inverse $G'_k(x)^\dagger$ exists?

In order to answer the first question, we first need to define a path θ of length N as a mode string $\theta_1\theta_2\ldots\theta_N$ with values in $\{1,\ldots,m\}$, and Θ_N as the set of all m^N paths of length N. We also define the observability matrix $\mathcal{O}(\theta)$ of a path θ of length N as:

$$\mathcal{O}(\theta) \triangleq \begin{pmatrix} C(\theta_1) \\ \vdots \\ C(\theta_N)A^{N-1} \end{pmatrix}. \tag{103}$$

Letting $\theta^k \triangleq \theta_k \ldots \theta_{k+N_B-1}$, it is clear, from the definition of G_k in (95), that x satisfies $G_k(x) = 0$ if and only if there exists a path $\theta \in \Theta_{N_B}$ such that

$$\mathcal{O}(\theta)x = \mathcal{O}(\theta^k)x_k. \tag{104}$$

The set of solutions to the measurement equation $G_k(x) = 0$ is therefore the union over all paths $\theta \in \Theta_{N_B}$ of the solutions to (104). First, taking $\theta = \theta^k$ in (104), we see that $O(\theta^k)$ must be of full column rank, otherwise any state x in the affine subspace $x_k + \ker(\mathcal{O}(\theta^k))$ solves $G_k(x) = 0$. In this case, x_k is clearly not even an isolated root of G_k. Therefore, since we assume that the mode sequence is arbitrary, we require at least that $\rho(\mathcal{O}(\theta)) = n$ for any path θ of length N_B, or *pathwise observability with* *index at most* N_B, as defined in [5]. However, even under such an assumption, the measurement equation $G_k(x) = 0$ may have multiple solutions, albeit a finite set, namely at most one for each $\theta \neq \theta^k$. A full analysis of this discrete aspect is given in [6], where it has been shown that it is impossible to guarantee x_k to be the unique global solution of $G_k(x) = 0$ for all $x_k \in \mathbb{R}^n$ (actually, not even for all $x_k \in \mathbb{R}^n \backslash \{0\}$) and all paths θ^k, which is yet another reason why local convergence is all we can aim for in Section 5.5. In any case, the set of solutions to $G_k(x) = 0$ being at most finite under pathwise observability, pathwise observability is established as a necessary and sufficient condition for x_k to always be an isolated root of G_k.

We now turn to the second question. $G'_k(x)$ is difficult to analyze unless $x = x_k$, in which case

$$G'_k(x_k) = - \begin{pmatrix} \left(\prod_{j \neq \theta_k} (C(\theta_k) - C(j)) x_k \right) C(\theta_k) \\ \vdots \\ \left(\prod_{j \neq \theta_{k+N_B-1}} (C(\theta_{k+N_B-1}) - C(j)) A^{N_B-1} x_k \right) \\ \times C(\theta_{k+N_B-1}) A^{N_B-1} \end{pmatrix}. \tag{105}$$

Actually, all we will study is $G'_k(x_k)$, and we will deduce that $G'_k(x)$ has full rank for x close enough to x_k, by taking advantage of the smoothness of G'_k. By analogy to scalar polynomials, we define *non-degeneracy* as follows:

Definition 17 *x is a non-degenerate root of G_k if $G_k(x) = 0$ and $G'_k(x)$ has full column rank.*

Note that non-degeneracy implies that x is the unique solution of $G_k(x) = 0$ in a neighborhood of x. Before stating the main result of this section, we need to define the function $\mathcal{P}$ of a pair of paths θ^1 and θ^2 as follows:

$$\mathcal{P}(\theta^1, \theta^2) \triangleq \mathcal{O}(\theta^1) - \mathcal{O}(\theta^2), \tag{106}$$

and we make the following assumption for further analysis:

Assumption 1 *Given system (91) and block size N_B, assume that there exist two integers $N_1 \geq n$ and $N_2 \geq n$ such that $N_B = N_1 + N_2 - 1$ and*

1. *For any path $\theta \in \Theta_{N_B}$, every $N_1 \times n$ submatrix of $\mathcal{O}(\theta)$ has full rank.*

2. *For any pair of paths θ^1 and θ^2 in Θ_{N_B}, if $\theta_i^1 \neq \theta_i^2 \ \forall i \in \{1, \ldots, N_B\}$, then every $N_2 \times n$ submatrix of $\mathcal{P}(\theta^1, \theta^2)$ has full rank.* $\diamondsuit$

Although the existence of an integer N such that $\mathcal{O}(\theta)$ has full rank for any $\theta \in \Theta_N$ has been shown to be decidable in [5], the decidability of determining the existence of a block size N_B satisfying Assumption 1 is still an open question. We have:

Lemma 8 *If Assumption 1 is satisfied, then x_k is a non-degenerate root of G_k whenever $x_k \neq 0$.*

Proof: From now on, all norms are Euclidean or induced Euclidean. We need to show that σ, defined as

$$\sigma \triangleq \inf_{\|t\|=1} \|(G_k'(x_k))t\|, \tag{107}$$

is positive. Defining the path θ^k as $\theta^k \triangleq \theta_k \cdots \theta_{k+N_B-1}$, and

$$\begin{aligned} \rho &\triangleq \inf_{\|t\|=1} \min_{\phi \in \Phi} \|\phi(\mathcal{O}(\theta^k))t\|, \\ \psi &\triangleq \min_{u \in U} \max_{i=1,\ldots,N_2} \prod_{j \neq u(i)} |(C(\theta^k_{u(i)}) - C(j))A^{u(i)-1}x|, \end{aligned}$$

where Φ is the set of all functions ϕ that extract an $N_1 \times n$ submatrix from a matrix $\mathcal{O}(\theta)$, and where U is the set of all maps $u : \{1, \ldots, N_2\} \to \{1, \ldots, N_B\}$, and noting that $\sigma \geq \psi\rho$, it suffices to show that ρ and ψ are both positive. That $\rho > 0$ follows directly from condition 1. Now, suppose that $\psi = 0$. Then there exists $u \in U$ such that

$$\max_{i=1,\ldots,N_2} \prod_{j \neq u(i)} |(C(\theta^k_{u(i)}) - C(j))A^{u(i)-1}x_k| = 0, \tag{108}$$

which implies that

$$\prod_{j \neq u(i)} |(C(\theta^k_{u(i)}) - C(j))A^{u(i)-1}x_k| = 0, \tag{109}$$

$\forall i \in \{1, \ldots, N_2\}$, which contradicts condition 2 since $x_k \neq 0$.

5.5 Convergence

In this section, we show that the DAA-Newton observer (99-100) results in a local exponential observer for (91). We have:

Theorem 21 *Assume that system (91) satisfies Assumption 1 and that A is invertible. Then, whenever $x_0 \neq 0$, the DAA-Newton observer (99-100) results in a local exponential observer for (91).* $\Diamond$

We now embark on proving Theorem 21. We first prove an essential lemma in Section 5.5.1, before detailing the proof in Section 5.5.2. For the remainder of the chapter, recall that the norm $\|\cdot\|$ is assumed to be the Euclidean (or induced Euclidean) norm. $B(x, r)$ denotes the open ball of radius r centered around x. The p^{th} differential of a function G is written $G^{\{p\}}$, but we will sometimes write $G' = G^{\{1\}}$ and $G'' = G^{\{2\}}$.

5.5.1 Newton Observers

Consider the nonlinear system

$$\begin{aligned} x_{k+1} &= f(x_k) \\ G_k(x_k) &= 0, \end{aligned} \tag{110}$$

where the measurement map G_k is time-varying and square or overdetermined (i.e. $G_k : \mathbb{R}^n \to \mathbb{R}^N$, $N \geq n$), so that we can define a Newton observer for (110) as follows:

$$\hat{x}_k^- = f(\hat{x}_{k-1}) \tag{111}$$

$$\hat{x}_k = \mathcal{F}_k(\hat{x}_k^-), \tag{112}$$

where $\mathcal{F}_k$ is given by:

$$\mathcal{F}_k(x) \triangleq x - (G_k'(x))^\dagger G_k(x). \tag{113}$$

In [58], the observer (111-112) was shown to be locally convergent for time-invariant systems that are controlled-invariant with respect to a compact set. Here, we present an extension of that result to a class of time-varying, possibly unstable autonomous systems described by (110). Furthermore, the following result does not require the strong observability assumptions of [58]. In particular, in [58], the augmented measurement equation is assumed to have a unique global solution (i.e. x such that $G_k(x) = 0$), which, as we have established, is impossible to guarantee for all $x_k \in \mathbb{R}^n$ if G_k is defined as in (95). This global observability condition therefore needs to be relaxed to a local observability condition. Moreover, the observability rank condition also has to be relaxed in a similar way. The following lemma incorporates these modifications. As can be anticipated, we will prove Theorem 21 by showing that the system in (98) satisfies the requirements of the following lemma, whose proof is given in Appendix B:

Lemma 9 *Consider the system in (110). First, assume that f and G_k, $k \geq 0$, are in $\mathcal{C}^3(\mathbb{R}^n)$, and that f is globally L-Lipschitz (i.e. for all x, y in $\mathbb{R}^n$, $\|f(x) - f(y)\| \leq L\|x - y\|$). Furthermore, assume that given $x_0 \in \mathbb{R}^n$, there exists a sequence R_k of subsets of $\mathbb{R}^n$ such that:*

1. *$x_k \in R_k$, $k \geq 0$,*
2. *defining d_k as $\operatorname{dist}(x_k, R_k^c)$, where R_k^c is the complement of R_k in $\mathbb{R}^n$, there exists $\beta > 0$ such that $d_{k+1} \geq \beta d_k > 0$,*
3. *and finally,*

   (a) $\exists\, g_p > 0$, $\gamma_p > 0$ such that $\sup_{x \in R_k} \|G_k^{\{p\}}(x)\| \leq g_p \gamma_p^k$, $p \in \{1, 2, 3\}$,

(b) $\exists\, g_\dagger > 0,\ \gamma_\dagger > 0$ *such that* $\|(G_k'(x_k))^\dagger\| \le g_\dagger \gamma_\dagger^k$.

Then there exist $c > 0$ *and* $\nu > 0$ *such that* $\frac{1}{2}\sup_{x\in X_k}\|\mathcal{F}_k''(x)\| \le c\nu^k$*, where* $X_k = \{x \in R_k \mid \|(G_k'(x))^\dagger\| \le 2g_\dagger\gamma_\dagger^k\}$*, and, moreover, the observer given by (111-112) results in a local exponential observer for (110), in the sense that if* $\hat{x}_0^-$ *satisfies:*

$$\begin{aligned} \|\hat{x}_0^- - x_0\| &\le \delta, \quad \textit{then} && (114)\\ \|\hat{x}_{k+1} - x_{k+1}\| &\le \alpha\|\hat{x}_k - x_k\|, && (115) \end{aligned}$$

for all $k \ge 0$*, whenever* α *and* δ *satisfy:*

- $0 < \alpha \le \min\left\{\beta, \frac{1}{\gamma_\dagger\gamma_2}, \frac{1}{\gamma_\dagger^2\gamma_1\gamma_2}, \frac{\beta}{\gamma_\dagger\gamma_1}, \frac{1}{\nu}\right\}$,
- $0 \le \delta < \min\left\{\frac{d_0}{2}, \frac{1}{4g_\dagger g_2}, \frac{1}{8g_\dagger^2 g_1 g_2}, \frac{d_0}{4g_\dagger g_1}, \frac{\alpha}{cL}\right\}$. $\Diamond$

Note that our definition of a local exponential observer does not imply that the rate α and radius δ of convergence are uniform over the entire state space. In other words, α and δ in Lemma 9 depend on x_0.

5.5.2 Proof of Theorem 21

The proof lies in showing that the system in (98) satisfies the requirements for Lemma 9, under the assumptions of Theorem 21. First of all, the dynamics being linear and G_k being polynomial in the state, they are both in $\mathcal{C}^3(\mathbb{R}^n)$. Moreover, A being invertible, there exist $l > 0$ and $L > 0$ such that

$$l\|x\| \le \|Ax\| \le L\|x\|, \tag{116}$$

$\forall x \in \mathbb{R}^n$. This implies that the dynamics is L-Lipschitz.

Next, since $x_0 \ne 0$, there exist $r_0 > 0$, $r_0' > 0$ such that $r_0 < \|x_0\| < r_0'$. Letting $r_k = r_0 l^k$ and $r_k' = r_0' L^k$, we get $x_k \in R_k$, $k \ge 0$, where $R_k \triangleq \{x \in \mathbb{R}^n \mid r_k < \|x\| < r_k'\}$. Clearly, $d_{k+1} \ge l d_k > 0$. It now remains to prove that conditions 3.(a) and 3.(b) in Lemma 9 are met.

At this point, we need to integrate the fact that the DAA-Newton observer converges for arbitrary mode sequences $\{\theta_k\}_{k=0}^{\infty}$. We thus define the parameterized function $\mathcal{G}_{\theta,x^*}$ as follows ($|\theta| = N$):

$$\mathcal{G}_{\theta,x^*}(x) \triangleq \begin{pmatrix} \displaystyle\prod_{i=1}^{m}(C(\theta_1)x^* - C(i)x) \\ \vdots \\ \displaystyle\prod_{i=1}^{m}(C(\theta_N)A^{N-1}x^* - C(i)A^{N-1}x) \end{pmatrix}, \tag{117}$$

and we note that, letting $\theta^k \triangleq \theta_k \cdots \theta_{k+N_B-1}$, we get $G_k(x) = \mathcal{G}_{\theta^k,x_k}(x)$, which implies that, for $p \in \{1,2,3\}$,

$$\sup_{x\in R_k} \|G_k^{\{p\}}(x)\| \leq \max_{\theta\in\Theta_{N_B}} \sup_{x^*\in R_k} \sup_{x\in R_k} \|\mathcal{G}_{\theta,x^*}^{\{p\}}(x)\|, \tag{118}$$

and we can therefore focus on bounding the right-hand side of (118) in proving that condition 3.(a) is met for any $\{\theta_k\}_{k=0}^{\infty}$. Since $r'_k = r'_0 L^k$, and since $\mathcal{G}_{\theta,x^*}^{\{p\}}(x)$ is polynomial in x^* and x, it is straightforward to show that there exist $g_p > 0$ and $\gamma_p > 0$, $p \in \{1,2,3\}$, such that

$$\max_{\theta\in\Theta_{N_B}} \sup_{\|x^*\|\leq r'_k} \sup_{\|x\|\leq r'_k} \|\mathcal{G}_{\theta,x^*}^{\{p\}}(x)\| \leq g_p \gamma_p^k, \tag{119}$$

for $k \geq 0$, and since $R_k \subset \{x \mid \|x\| \leq r'_k\}$, we get

$$\sup_{x\in R_k} \|G_k^{\{p\}}(x)\| \leq g_p \gamma_p^k. \tag{120}$$

As for condition 3.(b), we have

$$\|(G'_k(x_k))^\dagger\| \leq \max_{\theta\in\Theta_{N_B}} \sup_{x\in R_k} \|(\mathcal{G}'_{\theta,x}(x))^\dagger\|. \tag{121}$$

Lemma 8 tells us that $(\mathcal{G}'_{\theta,x}(x))^\dagger$ is defined, and therefore continuous, over the entire unit sphere. Therefore, since the unit sphere in $\mathbb{R}^n$ is compact, we get that

$$\mathcal{H} \triangleq \max_{\theta\in\Theta_{N_B}} \sup_{\|x\|=1} \|(\mathcal{G}'_{\theta,x}(x))^\dagger\| \tag{122}$$

is finite. Consequently, observing that $\|(\mathcal{G}'_{\theta,rx}(rx))^{\dagger}\| = \frac{1}{r_k^{m-1}}\|(\mathcal{G}'_{\theta,x}(x))^{\dagger}\|$ for any $r \in \mathbb{R}^*$,

$$\max_{\theta\in\Theta_{N_B}} \sup_{x\in R_k} \|(\mathcal{G}'_{\theta,x}(x))^{\dagger}\| \quad = \quad \frac{\mathcal{H}}{r_k^{m-1}}, \tag{123}$$

hence 3.(b) with $g_{\dagger} = \frac{1}{\mathcal{H}r_0^{m-1}}$ and $\gamma_{\dagger} = \frac{1}{l^{m-1}}$. $\square$

5.6 Numerical Results

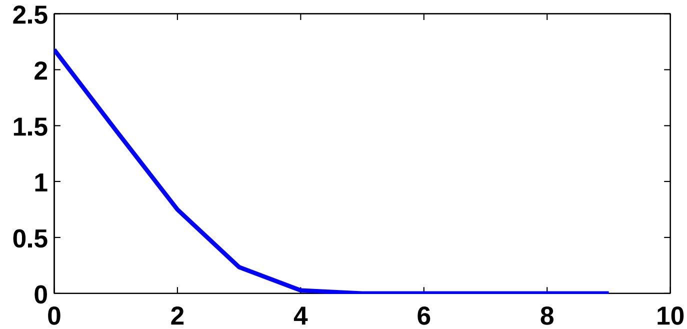
In this section, we evaluate the performance of the DAA-Newton observer (99-100) by numerical simulation. Let us consider system (91), with:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \begin{cases} C(1) & = & (1 \quad 0) \\ C(2) & = & (2 \quad 3), \end{cases} \tag{124}$$

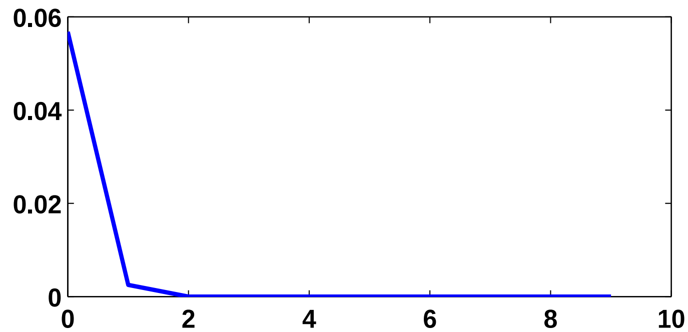
which satisfies the assumptions of Theorem 21 (with $N_1 = 2$, $N_2 = 2$ and $N_B = 3$). The radius of convergence is evaluated by numerical simulation to be $\simeq 0.4$ at $x_0 = (1,1)^T$. In Figure 1, the observer error $\|\hat{x}_k - x_k\|$ is plotted versus time for $\hat{x}_0^- = (0.7, 0.7)^T$ and for three different mode sequences θ, demonstrating the stability of the observer.

5.7 Conclusions

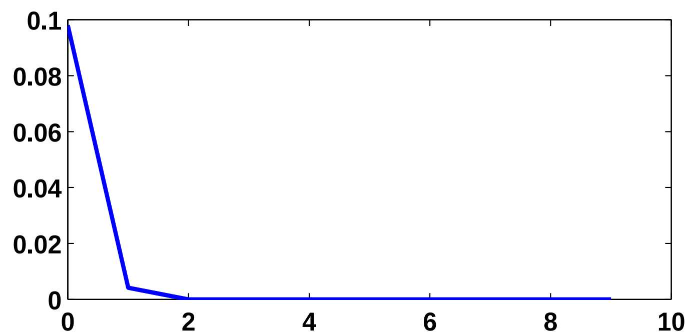
An observer design approach is presented for linear discrete-time systems with randomly-switching measurement equations. It is shown to produce local exponential observers, and numerical simulations support the soundness of the approach. Given the current results, the approach is being evaluated on more general systems, such as nonautonomous and multi-output systems. Moreover, using the DAA to design estimators for noisy systems is currently under investigation.



(a) $\theta = 1 \cdot 1 \cdot 1 \cdot 1 \cdots$



(b) $\theta = 2 \cdot 2 \cdot 2 \cdot 2 \cdots$



(c) $\theta = 1 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1 \cdots$

Figure 4: DAA-Newton observer error versus time for three different measurement mode sequences.

CHAPTER 6

CONCLUSIONS

6.1 Summary of Contributions

The contributions of this thesis can be summarized as follows.

- Pathwise observability has been shown to be decidable, and upper bounds on the indexes of pathwise observability have been given [5].

- It has been shown that for SLS's with switching in only the measurement equation, observability of each pair in continuous-time implies pathwise observability of the discretized system for almost any sampling period, which constitutes an extension of the Kalman-Bertram criterion to the switched case [7, 8].

- State observability and mode observability have been characterized in autonomous SLS's, and shown to be decidable [6].

- In the non-autonomous case, only mode observability has been characterized. It has also been shown that single-experiment strong mode observability and generic-experiment strong mode observability are equivalent [6].

- The Direct Algebraic Approach, a novel asymptotic observer design technique, has been proposed, and shown to be exponentially convergent under generic conditions [9, 10].

6.2 Future Work

We have identified the following subjects of future research.

- Only upper bounds $N(s,n,n)$ have been given on the maximum indexes $\mathcal{N}(s,n,n)$ of PWO. The bounds seem to be very conservative, and grow at an enormous rate, which makes PWO impossible to check in many cases given the current available computational power. It is therefore crucial to find the exact maximum indexes of PWO, or to at least try to reduce the upper bounds $N(s,n,n)$ to levels that allow for practical verification of PWO. Moreover, it is still unknown whether or not the maximum indexes of PWO depend on p, the dimension of the measurements.

- In non-autonomous SLS's, no decidability result is yet available. The decidability of SMO seems to be a challenging problem, and deserves to be studied, on the same basis that PWO has.

- So far, only *arbitrary* switching has been considered. The "existential" problems, where, e.g., one may wish to find an observable path, remain completely open.

- Mode observability under known states remains an open problem as well.

- The DAA has only been studied in the deterministic autonomous case. It would be interesting to find out to what extent it can be applied to non-autonomous and stochastic systems. In particular, by nature, the DAA would provide computationally efficient estimators for noisy switched linear systems, and it is therefore of great importance to compare it to existing hybrid estimators.

APPENDIX A

SOME GENERALIZED MATRIX INVERSION THEORY

We now present some definitions and results from matrix inversion theory (see, e.g., [21]). We first recall that $\mathcal{O}^{\{1\}}$ is a $\{1\}$-inverse of $\mathcal{O}$ if

$$\mathcal{O}\mathcal{O}^{\{1\}}\mathcal{O} = \mathcal{O}, \tag{125}$$

and that the (Moore-Penrose) pseudo-inverse of $\mathcal{O}$ is defined as

$$\mathcal{O}\mathcal{O}^{\dagger}\mathcal{O} = \mathcal{O},\ \mathcal{O}^{\dagger}\mathcal{O}\mathcal{O}^{\dagger} = \mathcal{O}^{\dagger},\ \mathcal{O}^{\dagger}\mathcal{O} = (\mathcal{O}^{\dagger}\mathcal{O})',\ \text{and}\ \mathcal{O}\mathcal{O}^{\dagger} \ = \ (\mathcal{O}\mathcal{O}^{\dagger})'. \tag{126}$$

Note that the pseudo-inverse $\mathcal{O}^{\dagger}$ of $\mathcal{O}$ always satisfies (125), and is therefore a $\{1\}$-inverse. If furthermore $\mathcal{O}$ has full column rank, then any $\{1\}$-inverse $\mathcal{O}^{\{1\}}$ of $\mathcal{O}$ is a left inverse of $\mathcal{O}$, in the sense that $\mathcal{O}^{\{1\}}\mathcal{O} = I$, the identity matrix. We next consider the following equation:

$$Y = \mathcal{O}x, \tag{127}$$

where $x \in \mathbb{R}^n$ and $Y \in \mathbb{R}^N$, and we examine the conditions on Y for (127) to have a solution in x, and how to compute that solution. Note that

$$\exists\ x \mid Y = \mathcal{O}x \Longleftrightarrow Y \in \Re(\mathcal{O}), \tag{128}$$

which is why we refer to the following test as the range inclusion test:

Proposition 8 (Range Inclusion Test) *If $\mathcal{O}^{\{1\}}$ is a $\{1\}$-inverse of $\mathcal{O}$, then*

$$Y \in \Re(\mathcal{O}) \Leftrightarrow (\mathcal{O}\mathcal{O}^{\{1\}} - I)Y = 0. \tag{129}$$

Proof:

$\Leftarrow$ Let $x = \mathcal{O}^{\{1\}}Y$. Then $Y = \mathcal{O}x$, which concludes the proof.

$\Rightarrow$ We have $Y \in \Re(\mathcal{O}) \;\Rightarrow\; \exists\, x$ s.t. $Y = \mathcal{O}x$. By definition of a left inverse, we have that $\mathcal{O}\mathcal{O}^{\{1\}}\mathcal{O}x = \mathcal{O}x$, which implies that $\mathcal{O}\mathcal{O}^{\{1\}}Y = Y$, which concludes the proof. $\square$

In words, equation (127) has a solution if and only if $(\mathcal{O}\mathcal{O}^{\{1\}} - I)Y = 0$ holds for some $\{1\}$-inverse (it then holds for *any* $\{1\}$-inverse). Note that if (127) admits a solution, then $x = \mathcal{O}^{\{1\}}Y$ is a solution to (127) for any $\{1\}$-inverse $\mathcal{O}^{\{1\}}$ of $\mathcal{O}$.

APPENDIX B

NEWTON OBSERVERS

B.1 A Classic Result

We first state the following standard result (adapted from [54, pp 279-281 & p 309]), which establishes the convergence of Newton's method:

Theorem 22 *Let G be a mapping from $\mathbb{R}^n$ to $\mathbb{R}^N$, where $N \geq n$, and assume that G is three times continuously differentiable. Assume further that:*

1. *There is a point $x_1 \in X$ such that $(G'(x_1))^\dagger$ exists with $\|(G'(x_1))^\dagger\| \leq \beta$ and $\|(G'(x_1))^\dagger G(x_1)\| \leq \eta$.*

2. *There exists $r \geq 2\eta$ such that $\sup_{x \in T} \|G''(x)\| \leq K$, where $T = B(x_1, r)$.*

3. *The constant $h = \beta\eta K$ satisfies $h < \frac{1}{2}$.*

Then the sequence $x_{n+1} = \mathcal{F}(x_n) \triangleq x_n - (G'(x_n))^\dagger G(x_n)$ of successive approximations generated by Newton's method exists for all $n \geq 1$, remains in T, and converges to a solution of $G(x) = 0$. Moreover, the rate of convergence is given by

$$\|x_{n+1} - x^*\| \leq \mu \|x_n - x^*\|^2, \tag{130}$$

where $\mu = \frac{1}{2} \sup\limits_{x \in T} \|\mathcal{F}''(x)\|$. ◊

B.2 Proof of Lemma 9

We now establish the lemma.

First, $\|\mathcal{F}_k''(x)\|$ needs to be adequately bounded. Schematically, note that for scalar $\mathcal{F}$, G and x, we have

$$\mathcal{F}'' = \frac{(G')^3 G'' - 2GG'(G'')^2 + G(G')^2 G'''}{(G')^4}.$$

This shows that $\mathcal{F}''_k(x)$ is polynomial in $G_k^{\{p\}}(x)$, $p \in \{1,2,3\}$, and $(G'_k(x))^\dagger$. Since these terms are bounded by exponentials over X_k, it is straightforward to bound the polynomial by an exponential, finding $c > 0$ and $\nu > 0$ such that $\frac{1}{2} \sup_{x \in X_k} \|\mathcal{F}''_k(x)\| \leq c\nu^k$, $k \geq 0$.

We now show by induction on k that

$$\|\hat{x}_k - x_k\| \;\leq\; \frac{\alpha}{L}\|\hat{x}_k^- - x_k\|, \qquad k \geq 0, \tag{131}$$

and note that (131), combined with the fact that f is globally L-Lipschitz, yields:

$$\|\hat{x}_{k+1} - x_{k+1}\| \;\leq\; \alpha\|\hat{x}_k - x_k\|, \qquad k \geq 0, \tag{132}$$

which establishes (115). Note that we also get $\|\hat{x}_{k+1}^- - x_{k+1}\| \leq \alpha\|\hat{x}_k^- - x_k\|$.

Equation (131) for $k = 0$ is a direct consequence of Lemma 10 and of (114).

Now, assume that (131) is true up to time $k = K - 1$, or in other words that $\|\hat{x}_k - x_k\| \leq \frac{\alpha}{L}\|\hat{x}_k^- - x_k\|$ for $0 \leq k \leq K-1$. Since f is globally L-Lipschitz, we furthermore have that $\|\hat{x}_{k+1}^- - x_{k+1}\| \leq L\|\hat{x}_k - x_k\|$ for $0 \leq k \leq K-1$. Combining these last two facts, we get

$$\|\hat{x}_K^- - x_K\| \leq \alpha^K\|\hat{x}_0^- - x_0\| \leq \alpha^K\delta, \tag{133}$$

which, again by Lemma 10, establishes (131) for $k = K$. $\square$

Lemma 10 *If* $\|\hat{x}_k^- - x_k\| \leq \alpha^k\delta$, *then* $\|\hat{x}_k - x_k\| \leq \frac{\alpha}{L}\|\hat{x}_k^- - x_k\|$. $\diamondsuit$

Proof: We first define, for $k \geq 0$:

- $\beta_k = \|(G'_k(\hat{x}_k^-))^\dagger\|$,
- $\eta_k = \|(G'_k(\hat{x}_k^-))^\dagger G_k(\hat{x}_k^-)\|$,
- $h_k = \beta_k\eta_k g_2\gamma_2^k$,
- $\rho_k = \min\left\{d_k, \frac{1}{2g_\dagger\gamma_\dagger^k g_2\gamma_2^k}\right\}$, and $S_k = B(x_k, \rho_k)$,

- $\mu_k = \frac{1}{2}\sup_{T_k}\|\mathcal{F}''_k(x)\|$, where $T_k = B(\hat{x}_k^-, 2\eta_k)$, and we note that $\mu_k \le c\nu^k$ if $T_k \subset X_k$.

We next note that since $\delta < \min\left\{\frac{d_0}{2}, \frac{1}{4g_\dagger g_2}\right\}$ and $\alpha \le \min\left\{\beta, \frac{1}{\gamma_\dagger\gamma_2}\right\}$, we have that $\alpha^k\delta < \frac{\rho_k}{2}$, and that $\hat{x}_k^- \in S_k$. Moreover, $S_k \subset R_k$ (with R_k given in Lemma 9) because $\rho_k \le d_k$. Therefore, by Lemma 11:

$$\sup_{x\in S_k}\|(G'_k(x))^\dagger\| \le 2g_\dagger\gamma_\dagger^k. \tag{134}$$

Thus, since $\hat{x}_k^- \in S_k \subset R_k$, $\|(G'_k(\hat{x}_k^-))^\dagger\| \le g_\dagger\gamma_\dagger^k$, which implies that $\beta_k \le g_\dagger\gamma_\dagger^k$ and $\eta_k \le \|(G'_k(\hat{x}_k^-))^\dagger\| \cdot \|G_k(\hat{x}_k^-)\| \le g_\dagger\gamma_\dagger^k\|G_k(\hat{x}_k^-)\| \le g_\dagger\gamma_\dagger^k g_1\gamma_1^k\|\hat{x}_k^- - x_k\| \le g_\dagger\gamma_\dagger^k g_1\gamma_1^k\alpha^k\delta$. Therefore, $h_k = \beta_k\eta_k g_2\gamma_2^k \le \delta g_\dagger^2 g_1 g_2\alpha^k\gamma_\dagger^{2k}\gamma_1^k\gamma_2^k$, and since $\delta < \frac{1}{8g_\dagger^2 g_1 g_2} < \frac{1}{2g_1 g_\dagger^2 g_2}$ and $\alpha \le \frac{1}{\gamma_\dagger^2\gamma_1\gamma_2}$, we get

$$h_k < \frac{1}{2}. \tag{135}$$

We now consider the open ball T_k. From $\delta < \min\left\{\frac{1}{8g_\dagger^2 g_1 g_2}, \frac{d_0}{4g_\dagger g_1}\right\}$ and $\alpha \le \min\left\{\frac{1}{\gamma_\dagger^2\gamma_1\gamma_2}, \frac{\beta}{\gamma_\dagger\gamma_1}\right\}$, we get that $2\eta_k \le 2g_\dagger\gamma_\dagger^k g_1\gamma_1^k\alpha^k\delta < \frac{\rho_k}{2}$, which, given that $\alpha^k\delta < \frac{\rho_k}{2}$, implies that $T_k \subset S_k \subset R_k$. Therefore,

$$\sup_{x\in T_k}\|G''_k(x)\| \le g_2\gamma_2^k. \tag{136}$$

Finally, by virtue of Theorem 22 and of (135) and (136), Newton's method would converge to a solution of $G_k(x) = 0$ inside T_k. By virtue of Lemma 11 and of the fact that $\rho_k = \min\left\{d_k, \frac{1}{2g_\dagger\gamma_\dagger^k g_2\gamma_2^k}\right\}$, x_k is the unique solution of $G_k(x) = 0$ in S_k. Since $T_k \subset S_k$, Newton's method converges to x_k, and we get, from (130), that

$$\|\hat{x}_k - x_k\| \le \mu_k\|\hat{x}_k^- - x_k\|^2, \tag{137}$$

and since $T_k \subset X_k$ (thanks to the fact that $T_k \subset S_k \subset R_k$ and to (134)), we have that $\mu_k \le c\nu^k$, which, combined with $\delta < \frac{\alpha}{cL}$, $\alpha \le \frac{1}{\nu}$, and (137), implies that

$$\|\hat{x}_k - x_k\| \le c\nu^k\alpha^k\delta\|\hat{x}_k^- - x_k\| \le \frac{\alpha}{L}\|\hat{x}_k^- - x_k\|, \tag{138}$$

which completes the proof. $\square$

The following straightforward lemma is provided without proof:

Lemma 11 *Let $G : R \to \mathbb{R}^N$, where R is an open subset of $\mathbb{R}^n$ and $N \geq n$. Assume that $G \in \mathcal{C}^2(R)$, and that there exists $x^* \in R$ such that $G(x^*) = 0$. Assume further that there exist two positive scalars $g_\dagger$ and g_2 such that $\|(G'(x^*))^\dagger\| \leq g_\dagger$ and $\sup_{x\in R} \|(G''(x))\| \leq g_2$, and let $r \leq \min\left\{\text{dist}(x^*, \overline{R}), \frac{1}{2g_\dagger g_2}\right\}$. then*

$$\sup_{x\in B(x^*,r)} \|(G'(x))^\dagger\| \leq 2g_\dagger, \tag{139}$$

and moreover, x^* is the unique solution of $G(x) = 0$ in $B(x^*, r)$. $\Diamond$

Finally, we give a graphical interpretation of Lemma 9 in Figure 2.

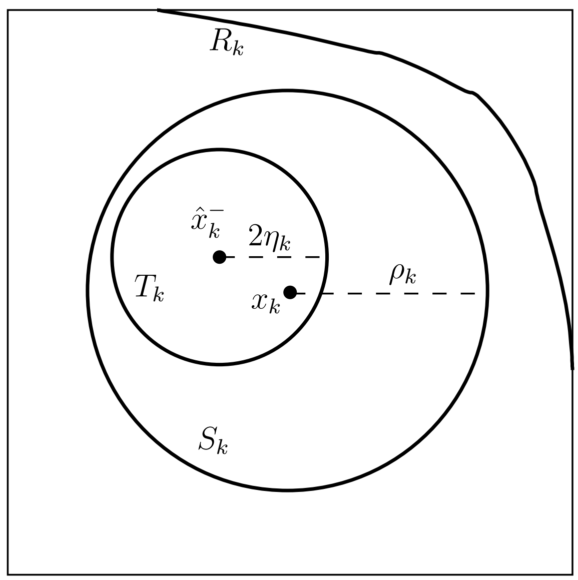


Figure 5: A graphical interpretation of Lemma 9. x_k is the unique solution of $G_k(x_k) = 0$ in S_k, and $\|(G_k(x))^\dagger\|$ is bounded over S_k. Moreover, $\hat{x}_k$ is guaranteed to be inside T_k with $\|\hat{x}_k - x_k\| \leq \frac{\alpha}{L}\|\hat{x}_k^- - x_k\|$.

REFERENCES

[1] Ackerson, G. A. and Fu, K. S., "On state estimation in switching environments," *IEEE Transactions on Automatic Control*, vol. 15, pp. 10–17, January 1970.

[2] Alcorta García, E., Seliger, R., and Frank, P. M., "Nonlinear decoupling approach to fault isolation in linear systems," in *Proceedings of the 1998 American Control Conference*, (Philadelphia, PA), pp. 2867–2871, June 1998.

[3] Alessandri, A. and Coletta, P., "Switching observers for continuous-time and discrete-time linear systems," in *Proceedings of the 2001 American Control Conference*, (Arlington, VA), June 2001.

[4] Alur, R., Henzinger, T. A., and Sontag, E. D., eds., *Hybrid Systems III: Verification and Control, Proceedings of the DIMACS/SYCON Workshop, October 22-25, 1995, Ruttgers University, New Brunswick, NJ, USA*, vol. 1066 of *Lecture Notes in Computer Science*, Springer, 1996.

[5] Babaali, M. and Egerstedt, M., "Pathwise observability and controllability are decidable," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, HW), pp. 5771–5776, December 2003.

[6] Babaali, M. and Egerstedt, M., "Observability of switched linear systems," Hybrid Systems: Computation and Control (R. Alur and G. Pappas, eds.), pp. 48–63, Springer, 2004.

[7] Babaali, M. and Egerstedt, M., "Pathwise observability for a class of switched linear systems: sufficient conditions and non-pathological sampling," *IEEE Transactions on Automatic Control, submitted*, January 2004.

[8] Babaali, M. and Egerstedt, M., "Pathwise observability through arithmetic progressions and non-pathological sampling," in *Proceedings of the 2004 American Control Conference*, (Boston, MA), June 2004.

[9] Babaali, M., Egerstedt, M., and Kamen, E. W., "An observer for linear systems with randomly-switching measurement equations," in *Proceedings of the 2003 American Control Conference*, (Denver, CO), pp. 1879–1884, June 2003.

[10] Babaali, M., Egerstedt, M., and Kamen, E. W., "A direct algebraic approach to observer design under switched measurement equations," *IEEE Transactions on Automatic Control, revised*, January 2004.

[11] Balluchi, A., Benvenuti, L., Di Benedetto, M. D., and Sangiovanni-Vincentelli, A. L., "Design of observers for hybrid systems," vol. 2289 of *Lect. Notes Comp. Sc.*, pp. 76–89, 2002.

[12] Balluchi, A., Benvenuti, L., Di Benedetto, M. D., and Sangiovanni-Vincentelli, A. L., "Observability for hybrid systems," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, HW), December 2003.

[13] BAR-SHALOM, Y. and LI, X. R., *Estimation and Tracking: Principles, Techniques, and Software*. Boston, MA: Artech House, 1993.

[14] BEMPORAD, A., FERRARI-TRECATE, G., and MORARI, M., "Observability and controllability of piecewise affine and hybrid systems," *IEEE Transactions on Automatic Control*, vol. 45, pp. 1864–1876, October 2000.

[15] BEMPORAD, A. and MORARI, M., "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, pp. 407–428, March 1999.

[16] BLOM, H. A. P., "An efficient filter for abruptly changing systems," in *Proceedings of the 23rd IEEE Conference on Decision and Control*, (Las Vegas, NV), December 1984.

[17] BLONDEL, V. D. and TSITSIKLIS, J. N., "Complexity of stability and controllability of elementary hybrid systems," *Automatica*, vol. 35, pp. 479–489, 1999.

[18] BLONDEL, V. D., THEYS, J., and VLADIMIROV, A. A., "An elementary counterexample to the finiteness conjecture," *SIAM Journal on Matrix Analysis and Applications*, vol. 24, pp. 963–970, 2003.

[19] BORKAR, V., MITTER, S., and TATIKONDA, S., "Markov control problems under communication constraints," *Communications in Information and Systems*, vol. 1, pp. 15–32, January 2001.

[20] BOUTAYEB, M. and DAROUACH, M., "Observers for linear time-varying systems," in *Proceedings of the 39th IEEE Conference on Decision and Control*, (Sydney, Australia), pp. 3183–3187, December 2000.

[21] CAMPBELL, S. L. and MEYER, C. D. J., *Generalized Inverses of Linear Transformations*. New York, NY: Dover, 1991.

[22] CHEN, T. and FRANCIS, B., *Optimal Sampled-Data Control Systems*. New York, NY: Springer, 1995.

[23] COLLINS, P. and VAN SCHUPPEN, J. H., "Observability of piecewise-affine hybrid systems," Hybrid Systems: Computation and Control, Springer-Verlag, 2004.

[24] COSTA, E. F. and DO VAL, J. B. R., "On the detectability and observability of discrete-time Markov jump linear systems," in *Proceedings of the 39th IEEE Conference on Decision and Control*, (Sydney, Australia), pp. 2355–2360, December 2000.

[25] DAUBECHIES, I. and LAGARIAS, J. C., "Sets of matrices all infinite products of which converge," *Linear Algebra and its Applications*, vol. 161, pp. 227–263, 1992.

[26] DE SANTIS, E., DI BENEDETTO, M. D., and POLA, G., "On observability and detectability of continuous-time linear switching systems," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, HW), December 2003.

[27] DEYST, J. J. and PRICE, C. F., "Conditions for asymptotic stability of the discrete minimum-variance linear estimator," *IEEE Transactions on Automatic Control*, vol. 13, pp. 702–705, December 1968.

[28] Doucet, A., Gordon, N. J., and Krishnamurty, V., "Particle filter for state estimation for jump Markov linear systems," vol. 49, pp. 613–624, March 2001.

[29] Elia, N., "Feedback stabilization in the presence of fading channels," in *Proceedings of the 2003 American Control Conference*, (Denver, CO), June 2003.

[30] Ezzine, J. and Haddad, A. H., "Controllability and observability of hybrid systems," *International Journal of Control*, vol. 49, no. 6, pp. 2045–2055, 1989.

[31] Ferrari-Trecate, G. and Gati, M., "Computation of observability regions for discrete-time hybrid systems," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, HW), December 2003.

[32] Fleming, W. H., *Report of the Panel on Future Directions in Control Theory: a mathematical perspective*. Philadelphia, PA: SIAM, 1989.

[33] Ge, S. S., Sun, Z., and Lee, T. H., "Reachability and controllability of switched linear discrete-time systems," *IEEE Transactions on Automatic Control*, vol. 46, pp. 1437–1441, September 2001.

[34] Gowers, W. T., "Fourier analysis and Szemerédi's theorem," *Doc.Math.J.DMV Extra Volume ICM I*, pp. 617–629, 1998.

[35] Graham, R. L., Rothschild, B. L., and Spencer, J. H., *Ramsey Theory*. New York: Wiley, 1990.

[36] Gurvits, L., "Stability of discrete linear inclusion," *Linear Algebra and its Applications*, vol. 231, pp. 47–85, 1995.

[37] Hwang, I., Balakrishnan, H., and Tomlin, C., "Observability criteria and estimator design for stochastic linear hybrid systems," in *Proceedings of the IEE European Control Conference*, (Cambridge, UK), September 2003.

[38] Ji, Y. and Chizeck, H., "Controllability, observability and discrete-time jump linear quadratic control," *International Journal of Control*, vol. 48, no. 2, pp. 481–498, 1988.

[39] Kabamba, P. T., "Control of linear systems using generalized sampled-data hold functions," *IEEE Transactions on Automatic Control*, vol. 32, pp. 772–783, 1987.

[40] Kailath, T., *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall, 1980.

[41] Kalman, R., Ho, B. L., and Narendra, N., "Controllability of linear dynamical systems," *Contributions to differential Equations*, vol. 1, pp. 198–213, 1963.

[42] Kalman, R. E., "Mathematical description of linear dynamical systems," *SIAM Journal of Control and Optimization*, vol. 1, pp. 152–192, 1963.

[43] Kamen, E. W., "Multiple target tracking based on symmetric measurement equations," *IEEE Transactions on Automatic Control*, vol. 37, pp. 371–374, March 1992.

[44] Kamen, E. W., "Block-form observers for linear time-varying discrete-time systems," in *Proceedings of the 32nd IEEE Conference on Decision and Control*, (San Antonio, TX), pp. 355–356, December 1993.

[45] Kamen, E. W., “Reduction of randomly-switching measurement equations to the deterministic case,” in *Proc. Conf. Info. Sc. & Syst*, John Hopkins University, 1997.

[46] Kreisselmeier, G., “On sampling without loss of observability/controllability,” *IEEE Transactions on Automatic Control*, vol. 44, pp. 1021–1025, May 1999.

[47] Kwakernaak, H. and Sivan, R., *Linear Optimal Control Systems*. New York: Wiley-Interscience, 1972.

[48] Lagarias, J. C. and Wang, Y., “The finiteness conjecture for the generalized spectral radius of a set of matrices,” *Linear Algebra and its Applications*, vol. 214, pp. 17–42, 1995.

[49] Li, X. R., “Hybrid estimation techniques,” *Control & Dynamical Systems*, vol. 76, pp. 213–287, 1996.

[50] Liberzon, D., Hespanha, J. P., and Morse, A. S., “Stability of switched systems: a lie-algebraic condition,” *Systems & Control Letters*, vol. 37, pp. 117–122, June 1999.

[51] Liberzon, D. and Morse, A. S., “Basic problems in stability and design of switched systems,” *IEEE Control Systems Magazine*, vol. 37, pp. 59–70, October 1999.

[52] Lidskii, E. A., “Optimal control of systems with random properties,” *Applied Mathematics & Mechanics*, vol. 27, pp. 33–45, 1963.

[53] Luenberger, D. G., “Observers for multivariable systems,” *IEEE Transactions on Automatic Control*, vol. 11, pp. 190–197, 1966.

[54] Luenberger, D. G., *Optimization by Vector Space Methods*. New York, NY: John Wiley & Sons, Inc, 1969.

[55] Mariton, M., “Stochastic observability of linear systems with markovian jumps,” in *Proceedings of the 25th IEEE Conference on Decision and Control*, (Athens, Greece), pp. 2208–2209, December 1986.

[56] Mariton, M., *Jump Linear Systems in Automatic Control*. New York, NY: Marcel Dekker, 1990.

[57] Middleton, R. and Freudenberg, J., “Non-pathological sampling for generalized sampled-data hold functions,” *Automatica*, vol. 31, no. 2, pp. 315–319, 1995.

[58] Moraal, P. E. and Grizzle, J. W., “Observer design for nonlinear systems with discrete-time measurements,” *IEEE Transactions on Automatic Control*, vol. 40, pp. 395–404, March 1995.

[59] Pattipati, K. R. and Sandell, Jr., N. S., “A unified view of state estimation in switching environments,” in *Proceedings of the 1983 American Control Conference*, pp. 458–465, 1983.

[60] Shelah, S., “Primitive recursive bounds for van der Waerden numbers,” *Journal of the American Mathematical Society*, vol. 1, pp. 683–697, 1988.

[61] Siegelman, H. T. and Sontag, E. D., “On the computational power of neural nets,” *Journal of Computation, Systems, and Science*, vol. 50, pp. 132–150, 1995.

[62] Smith, S. C. and Seiler, P., "Optimal pseudo-steady-state estimators for systems with markovian intermittent measurements," in *Proceedings of the 2002 American Control Conference*, (Anchorage, AK), pp. 3021–3027, May 2002.

[63] Sontag, E. D., "On the observability of polynomial systems, I: Finite-time problems," *SIAM Journal on Control and Optimization*, vol. 17, no. 1, pp. 139–151, 1979.

[64] Sontag, E. D., "Nonlinear regulation: The piecewise linear approach," *IEEE Transactions on Automatic Control*, vol. 26, pp. 346–358, April 1981.

[65] Sontag, E. D., "Real addition and the polynomial hierarchy," *Information Processing Letters*, vol. 20, pp. 115–120, 1985.

[66] Sontag, E. D., "Interconnected automata and linear systems: A theoretical framework in discrete-time," Hybrid Systems III: Verification and Control (R. Alur, T. Henzinger, and E. D. Sontag, eds.), pp. 436–448, Springer, 1996.

[67] Sworder, D. D., "Feedback control of a class of linear systems with jump parameters," *IEEE Transactions on Automatic Control*, vol. 14, pp. 9–14, February 1969.

[68] Tugnait, J. K., "Detection and estimation for abruptly changing systems," *Automatica*, vol. 18, no. 5, pp. 607–615, 1982.

[69] van der Waerden, B. L., "Beweis einer baudetschen vermutung," *Nieuw Archief voor Wiskunde*, vol. 15, pp. 212–216, 1927.

[70] Verriest, E. I., "Multi-mode system identification," in *Adaptive Control of Nonsmooth Dynamic Systems*, ch. 7, New York: Springer, 2001.

[71] Vidal, R., Chiuso, A., and Soatto, S., "Observability and identifiability of jump linear systems," in *Proceedings of the 41st IEEE Conference on Decision and Control*, (Las Vegas, NV), pp. 3614–3619, December 2002.

[72] Vidal, R., Chiuso, A., Soatto, S., and Sastry, S., "Observability of linear hybrid systems," Hybrid Systems: Computation and Control, Springer-Verlag, 2003.

[73] West, P. D. and Haddad, A. H., "On the observability of linear stochastic switching systems," in *Proceedings of the 1994 American Control Conference*, (Baltimore, MD), pp. 1846–1847, June 1994.

[74] Zhirabok, A. N. and Preobrazhenskaya, O. V., "Nonlinear methods for fault detection and isolation in linear systems," in *Proceedings of the 33rd IEEE Conference on Decision and Control*, (Lake Buena Vista, FL), pp. 3054–3055, December 1994.

VITA

Mohamed Babaali was born in 1975 in Blida, Algeria. He received the Diplôme d'ingénieur from Supélec in 2000, and the M.S. degree in Electrical and Computer Engineering from Georgia Tech in 2001. His research interests include signal processing and control theory.